

Corso FiXo / Master in comunicazione della Scienza
Università di Cagliari, 23-24 ottobre 2008

La comunicazione nel mondo scientifico

Francesca Di Donato didonato@sp.unipi.it
Università di Pisa <http://www.sp.unipi.it/hp/didonato>
(le slide del corso sono liberamente scaricabili da qui)

Programma del corso

Lezione 1. I motori di ricerca

- (1). La topologia del web
- (2). In pratica: Strumenti, regole, consigli
 - a. Ricerche di breve durata

Lezione 2. I motori di ricerca

- (2). In pratica: Strumenti, regole, consigli
 - b. Ricerche scientifiche (di lunga durata)

Lezione 3. Strumenti per selezionare e archiviare i risultati

- (1). Web semantico e web 2.0
- (2). Strumenti collaborativi per la ricerca

Lezione 4. Open Access

- (1). Storia e filosofia
- (2). Strumenti, problemi, prospettive

Bibliografia

A-L. Barabàsi, Link. La nuova scienza delle reti, Einaudi, Torino 2004.

Fravia, Searchlores. Advanced Internet searching strategies & advice. Resources for basic, advanced & deep web seekers, <http://www.searchlores.org> (pagine specifiche del sito sono indicate nelle singole slide).

T. Berners Lee, "The World Wide Web - Past, present, future. Exploring Universality" (2002), <http://www.w3.org/2002/04/Japan/Lecture.html>

Prima di metterci a cercare sul web dobbiamo conoscerne la **topologia**

Il web è un **grafo**, un insieme di **vertici (nodi)** uniti da **spigoli (link)**.



Grafo: i ponti di Königsberg

La sua architettura si basa su tre principi:

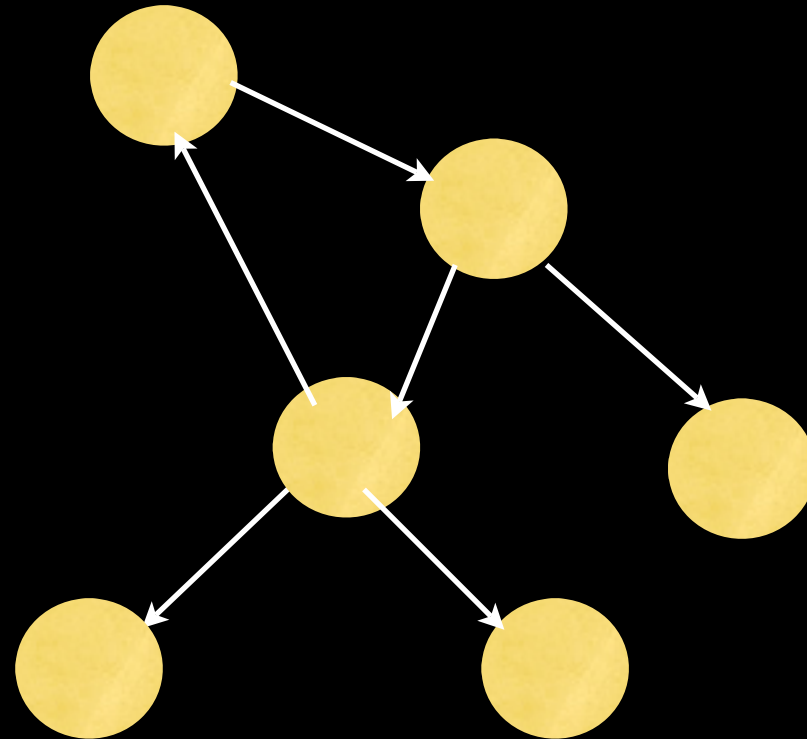
- 1) **URI/URL** (puntatori a nodi)
- 2) **HTTP**
- 3) **HTML** (link)

Un'importante proprietà dei grafi: **l'esistenza di un percorso da un nodo a un altro non dipende dalla nostra capacità di trovarlo ma è una proprietà del grafo.** (Eulero, 1736)

Il Web è una rete diretta,
complessa e a invarianza di scala

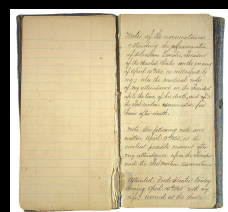
Cioè?

a) il web è un grafo orientato
(rete diretta)

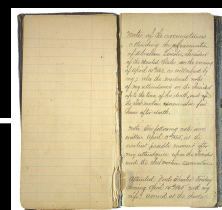


altri esempi?

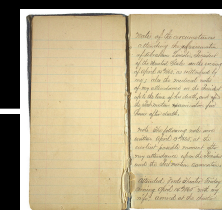
la rete della comunità scientifica (scritti e citazioni sono rispettivamente nodi e link)



Platone

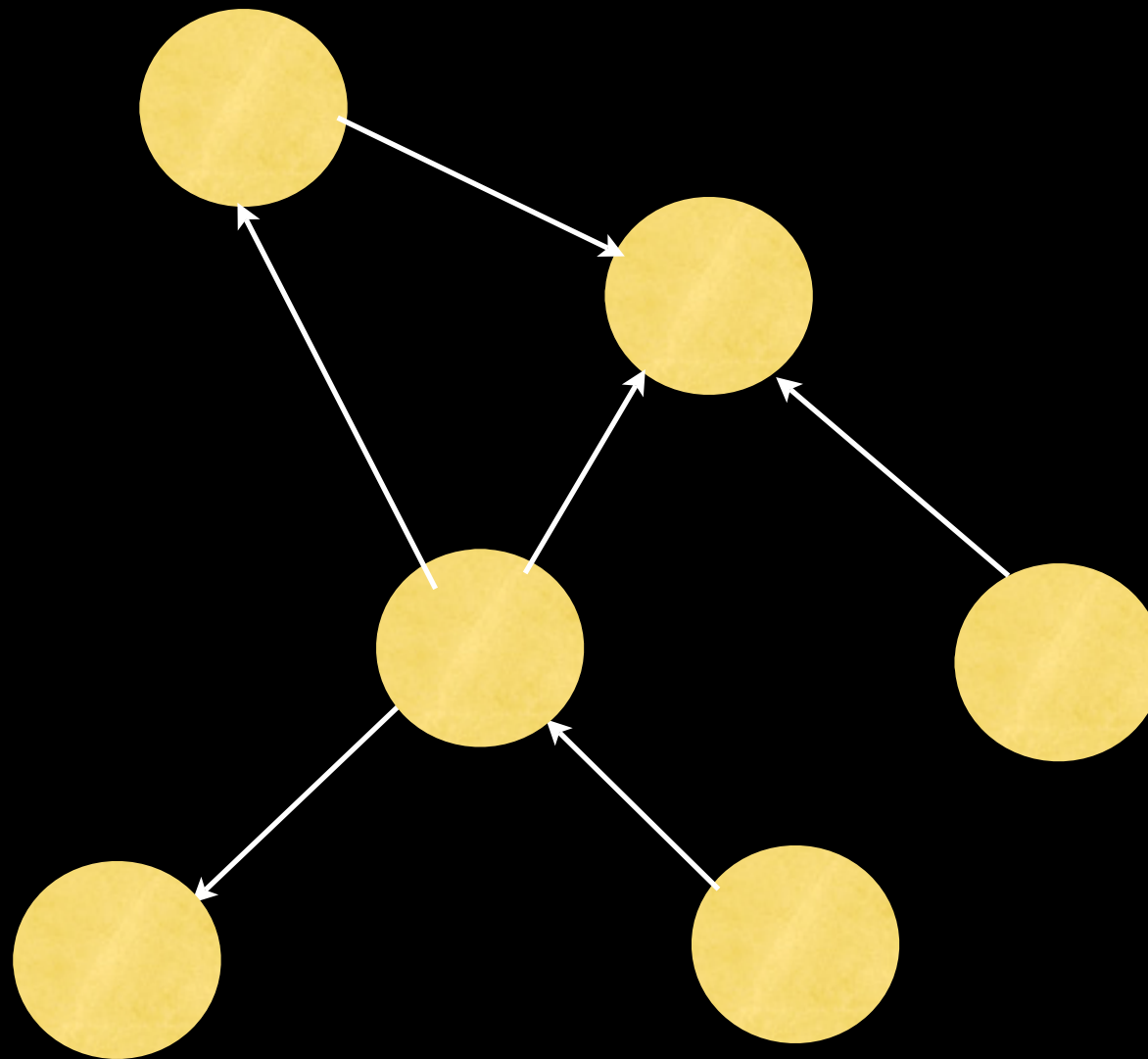


Kant



Di Donato

b) per far parte del grafo, **basta un link**
per ogni nodo



c) il web è un mondo piccolo

“Sei gradi di separazione” (Milgram): quanti contatti (link) sono necessari per connettere tra loro due individui (nodi) scelti a caso nella società?

Risposta: in media il numero minimo di intermediari necessari è 5,5. La società è una rete di 6 miliardi di nodi in cui la distanza media tra un nodo e l'altro non supera i 6 link.

Un mondo piccolo è una ragnatela molto fitta

anche la rete delle pubblicazioni - citazioni scientifiche è un mondo piccolo.

e il web?

“Supponiamo che i dati memorizzati dai computer di tutto il mondo siano connessi tra loro... Le informazioni più preziose del Cern e dell'intero pianeta sarebbero a disposizione mia e di chiunque altro. Esisterebbe un unico spazio globale dell'informazione.”

(Tim Berners-Lee, 1980)

2004: nello spazio globale dell'informazione i gradi di separazione sono 19.

d) presenza di connettori (hub)

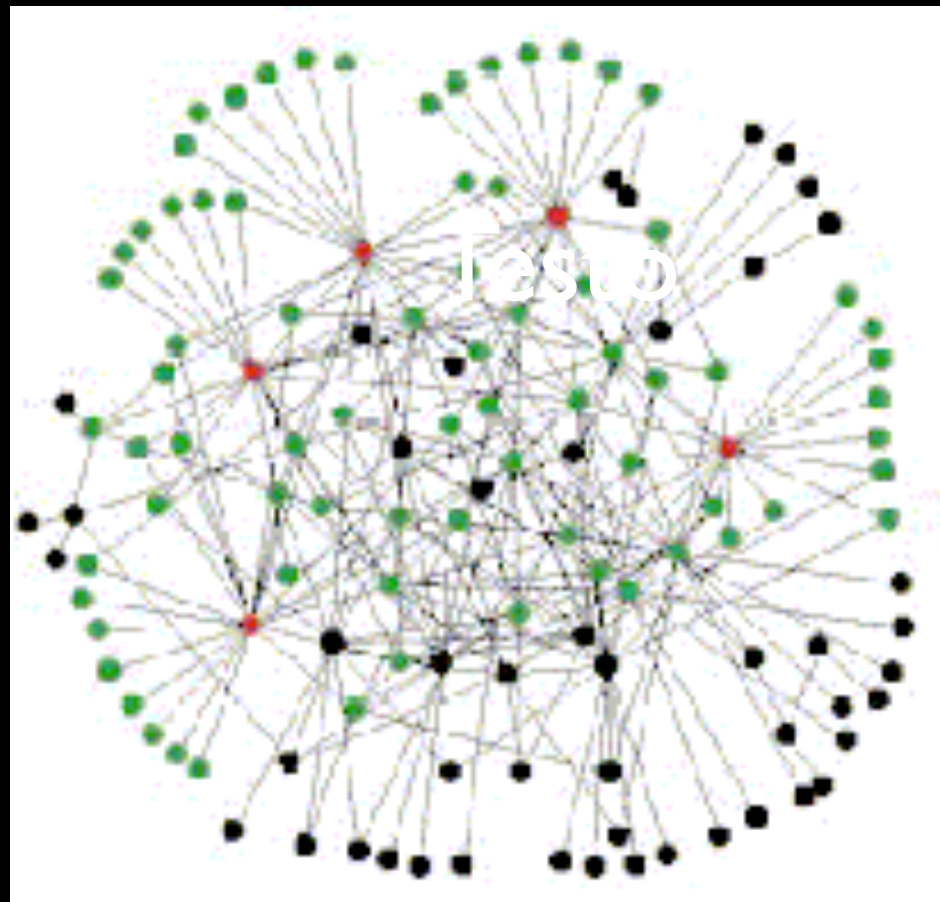
Disseminata in varie occupazioni, esiste una manciata di persone che possiede l'abilità davvero straordinaria di stringere un numero eccezionale di amicizie e conoscenze. Sono i connettori.

I connettori sono presenti in tutte le grandi reti complesse.

Esempi di connettori (hub) sul web?

e) il **numero di link** misura la visibilità di un nodo

L'architettura del WWW è dominata da pochissimi nodi altamente connessi (hub) che tengono insieme molti nodi poco connessi.



A scale-free network of 130 nodes generated by the scale-free model.

The five biggest nodes are shown in red, and they are in contact with 60% of nodes (green).

<http://www.nd.edu/~networks/Linked/>

NB. La presenza di hub rende queste reti forti agli attacchi (corollario: la rete si mette in crisi attaccando gli hub)

f) il principio 80/20

l'80% dei piselli viene dal 20% di baccelli;

l'80% dei profitti è prodotto dal 20% di impiegati;

l'80% di link parte dal 20% di pagine;

l'80% citazioni proviene dal 35% degli scritti;

etc.

g) il web è una **rete a invarianza di scala**

I nodi sono distribuiti secondo leggi di potenza che si basano su 2 principi:

1. **crescita** (nuovi nodi si aggiungono alla rete)
2. **collegamento preferenziale** (si preferisce aggiungere link verso pagine che hanno molti link).

h) il web è **un mondo frammentato...**

Il WWW è anche un mondo grande, perché è fatto di “**sottomondi**”. Ciò dipende dal fatto che è un grafo orientato (rete diretta).

Un'altra conseguenza di questa proprietà è che **non forma una singola rete omogenea.**

qual è allora la “forma” del web?

E' diviso in 4 continenti di dimensioni pressoché equivalenti

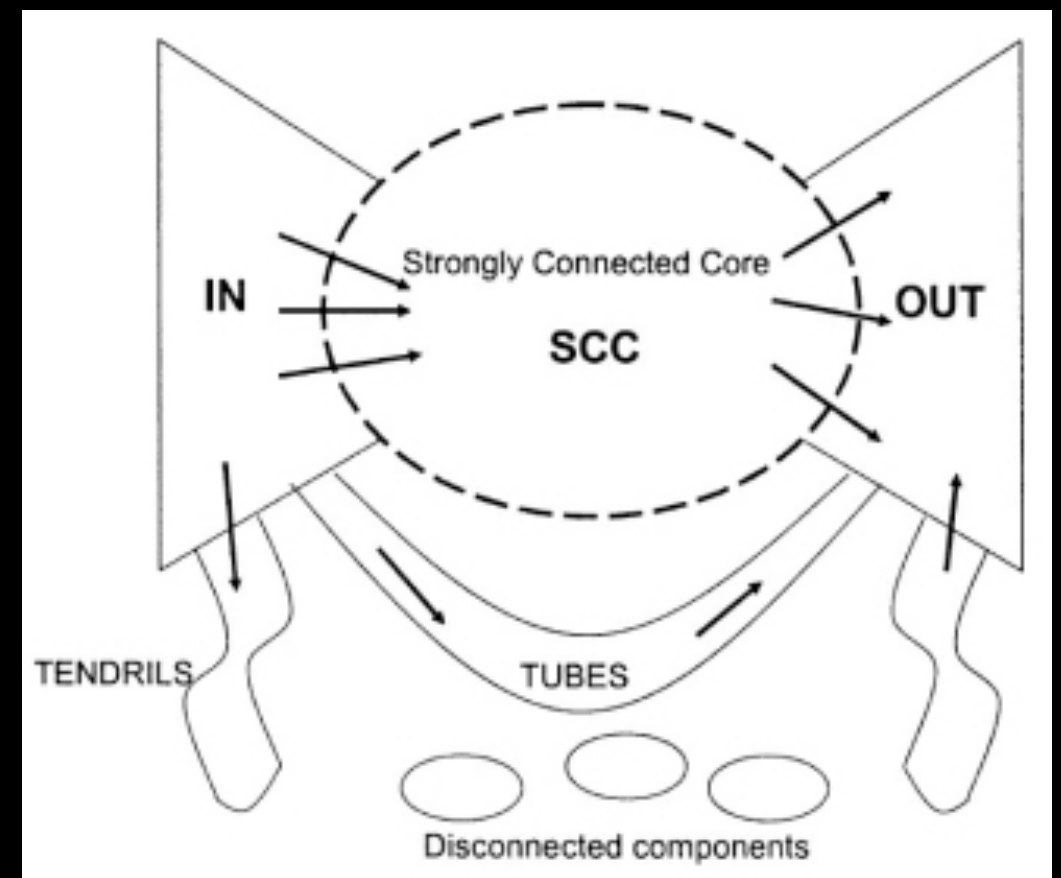
Il primo è il **corpo centrale (SCC)**. E' facile da navigare perché molto interconnesso, esiste sempre un percorso per muoversi al suo interno.

Il secondo, detto continente **IN** è più difficile da navigare perché ha collegamenti verso SCC ma non da SCC.

Il terzo continente (**OUT**) è raggiungibile da nodi del corpo centrale, ma non ci sono link per tornare indietro (è fatto di siti “a fondo chiuso”).

il quarto continente è fatto di **tentacoli**, **isole** separate e **tubi**, gruppi di pagine collegate tra loro ma non al corpo centrale.

<http://thenextweb.org/wp-content/uploads/2008/06/fragmented.jpg>



NB. Anche la rete delle
citazioni scientifiche è una
rete diretta e obbedisce alle
stesse regole

I motori di ricerca
mappano **solo il 50%** del
web, cioè il corpo centrale
e il continente OUT.

...il resto non è navigabile

Comincia la ricerca...

Sappiamo che il web è navigabile solo in parte. Ma come lo facciamo? Il nostro strumento principale sono i **motori di ricerca**.

Ci sono moltissimi motori, eppure usiamo quasi solo google (e lo facciamo male).

google ha una copertura molto ampia, ma non è necessariamente il migliore per la nostra ricerca. Perché? Dipende da che cosa cerchiamo.

Come si impara a cercare “bene”?

searchlores

Il sito dell'hacker finlandese Fravia:




















<http://www.searchlores.org>

Non è facile da usare, ma anche imparare ad orientarsi al suo interno è un esercizio per imparare a fare ricerca sul web;-)

Leggete il warning:

<http://www.searchlores.org/warning.htm>

I motori sono tanti....

Fravia's searching MAPA (<u>m</u> asks and <u>p</u> ages)  : has cache  : special page  : has NEAR <u>ss.ee. operators</u>						
Best s.e.	Ask Jeeves  (date)	MSNsearch 	Yahoo!    [YAHOO]	Google   [GOOGLE]		
Second Tier	Alexa 	Hotbot  [HOTBOT]	A 9 (google's )	Teoma/Ask  [TEOMA]	Fast  [FAST]	Alta!  Adva Simple
Useful s.e.	Wisenut	Baidu 	Gigablast 		IceRocket (webarchive)	Exalead (date & regexp) 
Graph s.e.	Kart00 (graph)	Touch (graph)	Ujiko (graph)	Dicy (cluster)	Mooter (cluster)	
Other				Lycos	Looksmart	Excite (ill)
Other	Entireweb	Wayback (past)	Factbite (ency)	dmoz (directory)	Furl (webarchive)	[FTPSEARCH]
@ PHP	[Our searching scrolls!]			[600 engines for next to nothing]		
@ fravia's	Targets	Local	Regional	Compound	Usenet	Accmail
@ fravia's	Live searches	Page Providers	Combing	Details	Databases	Allinones
@ fravia's	Images	Books	Laws	Files	Filez	Passwords
Bogus & crap				look.com	2020search	tygo.com

e devono essere **studiati e provati!!**

<http://www.searchlores.org/main.htm>

Usare gli operatori..

Qui una lista degli operatori dei principali motori di ricerca:

<http://www.searchlores.org/operators.htm>

Quelli di google, ad esempio, sono i seguenti:

site:

allintitle: (all of the query words in the title)

intitle: (that word in the title)

allinURL: (all of the query words in the URL)

inURL: (that word in the URL)

cache:

link:

related: (pages that are "similar" to a specified web page)

info: (google's info)

A. ricerca puntuale o di
breve periodo:

12 regole “auree”

(<http://www.searchlores.org/tips.htm>)

I. Fate mente locale sulla vostra query

a) pensate alle vostre domande (un suggerimento:
<http://www.searchlores.org/london/lea2tra.htm>)

Joe A. User's 7 questions

I collected the following queries on the morning of 5 february 2003, taking them at random from [Jeeves' peek](#), where you can see what people are searching for on jeeves eo ipso (many similar services are listed on the [ad hoc section](#) of my sites).

Here we go with 7 questions (by "Joe A. User", where the A stays for 'average') that will allow me to show you, to-day, some searching approaches --and caveats-- you may not be aware of.

Joe A. User's 7 questions		
Where can I find the flag of the country Ireland?	ubi	the real meaning of a query
Why is the sun yellow?	cur	knowing how to rephrase
Where can I find information about the chemical element chlorine?	quibus auxiliis	keeping the seeker's balance
How do I find the Italian national hymn?	quomodo	regional searching
When was the battle of Kadesh fought?	quando	evaluation lore (1)
Who is the father of modern biotechnology?	quis	evaluation lore (2)
What is the neurological condition acoustic neuroma?	quid	local and specialized engines

As you can see, all these questions have been presented in normal, plain English. Now let's have a look at the answers given by "ask jeeves", at the answers you should be able to find quickly, and at the possible approaches to find SIMILAR answers on the web. Let us see also whether the algorithms used by the various search engines are good enough as they are, are miserable, could be improved, should be ditched, or whatever.

b) pensate al tipo di risorse che vi servono per rispondere
usenet, blog o server ftp?
news, immagini, articoli, ecc.?

2. Usate varie risorse

Quale motore di ricerca è più utile per voi? Di nuovo: dipende da quello che state cercando.

E' importante conoscere le caratteristiche dei diversi motori di ricerca, le cui interrogazioni danno risultati diversi e non sovrapposti e che possono rispondere a diverse esigenze.

Una cosa è se cercate un articolo scientifico, un'altra se volete una poesia polacca di un secolo fa, un'altra ancora se cercate un MP3.....

3. Usate le minuscole

Anche se per la maggior parte dei motori di ricerca non fa più differenza, è sempre una buona regola scrivere le proprie query minuscole.

4. Usate la ricerca per sequenza esatta [“”]

ES: "searchengines" vi restituirà le pagine con l'esatto termine, "search engines" vi darà pagine con i due termini separati (da uno spazio o da un segno di interpunzione) mentre "search engines results" OR "searchengines results" OR "s.e. results" vi darà come risultato pagine con una di queste tre occorrenze.

(e inserite errori di ortografia volontari)

NB. A volte, inserire volutamente errori di battitura o ortografici può aiutare a trovare pagine che altrimenti sarebbero invisibili.

Per esempio "saerch engines" permette di trovare le pagine con lo stesso errore.

5. Usate [AND | & | +] e [AND NOT | | | -]

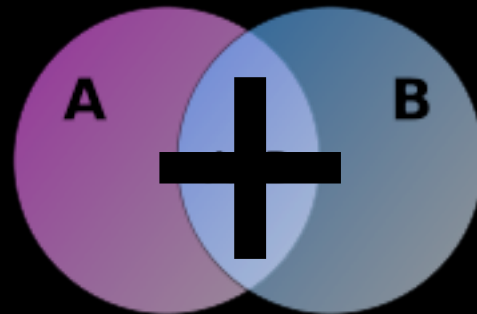
Affinate le vostre ricerche usando AND, & oppure +. Allo stesso modo, escludete alcune pagine usando AND NOT o | o semplicemente il segno meno [-].

Ad esempio: +"search engines" +hints +tips +techniques -tits -sex -"make money" è migliore rispetto a +"search engines" +hints +tips +techniques

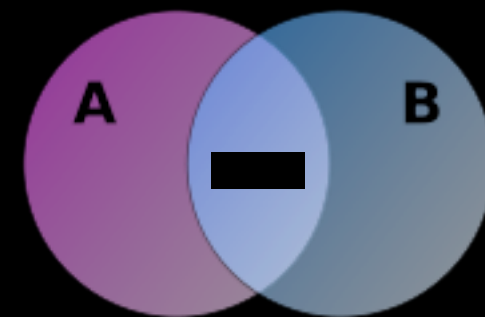
6. Sull'uso dei segni + e -

col segno + potreste perdere documenti che non hanno le parole che specificate: il risultato della vostra ricerca vi dà l'intersezione dei due insiemi.

per esempio, la ricerca "searching tips" +searchlores restituisce solo le pagine che abbiano entrambe le espressioni.



Lo stesso vale per il segno meno.



7. Sull'uso degli operatori booleani

L'uso degli operatori cambia da motore a motore. Alcuni, come Fast/Alltheweb, hanno box speciali per inserire pochi operatori (AND, OR and NOT).

In generale, è importante stare attenti nell'uso delle parentesi.

8. stringhe e parole (articoli e preposizioni)

E' conveniente eliminare dalla ricerca le parole che non sono particolarmente utili alla ricerca. Di queste fanno parte articoli e **preposizioni** (ad esempio, in inglese: the, and, a, to, of, in, i, is, that, it, on, you, this, for, but, with, are, have, be, at, or, as, was, so, if, out, not).

In generale, cercate di usare parole che sottolineino la peculiarità del vostro obiettivo. Le parole comuni, combinate attraverso operatori booleani, possono rivelarsi molto efficaci. Dovete identificare i concetti principali nel vostro argomento e definire sinonimi, spelling alternativi o forme varianti di parole per i concetti (una specie di mappa).

Ci sono motori (ad esempio Kartoo) che vi aiutano a espandere la vostra mappa.

8. stringhe e parole (articoli e preposizioni)

La scelta di parole o stringhe particolari vi aiuta ad affinare la vostra ricerca.

Un esempio: aggiungiamo la stringa "search strateg*", che ha un'elevata peculiarità, come parola chiave del titolo:

+ title:"search strateg*" +hints +tips

Incidentalmente: questo tipo di ricerca serve anche a scoprire i plagi:-)

9. Parole chiave (keywords) speciali

Osservate l'uso di una keyword (o operatore avanzato) nell'esempio precedente.

- Qui una breve lista delle principali keyword di altavista:

- * anchor:text
- * applet:class
- * domain:domainname to avoid commercial crap exclude with
-"domain:com"
- * host:name
- * image:filename
- * link:URLtext
- * text:text
- * title:text very useful for narrowing
- * url:text

Ogni motore di ricerca, lo ripetiamo, ha i suoi operatori avanzati, che è bene conoscere.

10. Usate l'asterisco [*]

Osservate anche l'uso dell'asterisco, nella query al punto 8:

`+ title:"search strateg*" +hints +tips`

di solito, dev'essere usato dopo almeno 3 caratteri dall'inizio del termine.

II. Usate motori di ricerca regionali e trucchi linguistici

Questo consiglio può essere molto utile a chi fa ricerca scientifica. E' utile tradurre le proprie “parole/stringhe chiave” nelle lingue conosciute e usare motori regionali. La profondità che possiamo raggiungere usando un motore coreano non può essere superata da nessuna ricerca su google.

12. Archivate le vostre ricerche

Durante le ricerche, perdersi è molto facile. Per seguire il proprio percorso, è importante annotare i risultati e usare alcuni strumenti. Ci sono i segnalibri dei browser, documenti di testo, EndNote (a pagamento), ma si può fare molto di più.

Come, lo vedremo nella terza lezione (qui le slide <<http://bfp.sp.unipi.it/~didonato/ovre/cagliari3.pdf>>).

Pausa...



Programma del corso

Lezione 1. I motori di ricerca

- (1). La topologia del web
- (2). In pratica: Strumenti, regole, consigli
 - a. Ricerche di breve durata

Lezione 2. I motori di ricerca

- (2). In pratica: Strumenti, regole, consigli
 - b. Ricerche scientifiche (di lunga durata)

Lezione 3. Strumenti per selezionare e archiviare i risultati

- (1). Web semantico e web 2.0
- (2). Strumenti collaborativi per la ricerca

Lezione 4. Open Access

- (1). Storia e filosofia
- (2). Strumenti, problemi, prospettive

il web è una rete di
biblioteche scientifiche:
come possiamo usarlo
per fare ricerca?

8

regole “auree”

([http://www.searchlores.org/
longtermsearching.htm](http://www.searchlores.org/longtermsearching.htm))

I. Sviluppare la vostra strategia di ricerca: scrivete il vostro piano

In pratica... formulate la query dividendo le sue componenti, annotando i vari termini (o concetti) e le loro definizioni e preparando una prima breve lista di sinonimi e/o sineddoche.

un esempio: immaginate di cercare modi e tecniche per capire qual è la vera lingua madre di un autore di una mail in inglese.

E' un concetto così vago e complesso da non poter essere descritto attraverso un insieme specifico di termini. In casi come questo, dovrete provare ad avvicinarvi al problema tramite tentativi che possano affinare la vostra ricerca. All'inizio, userete termini chiari e specifici, che poi modificherete nel corso della ricerca.

2. Stabilite una priorità tra i termini di ricerca e i concetti

Una volta stilata una lista di termini, dovete mettervi ordine.

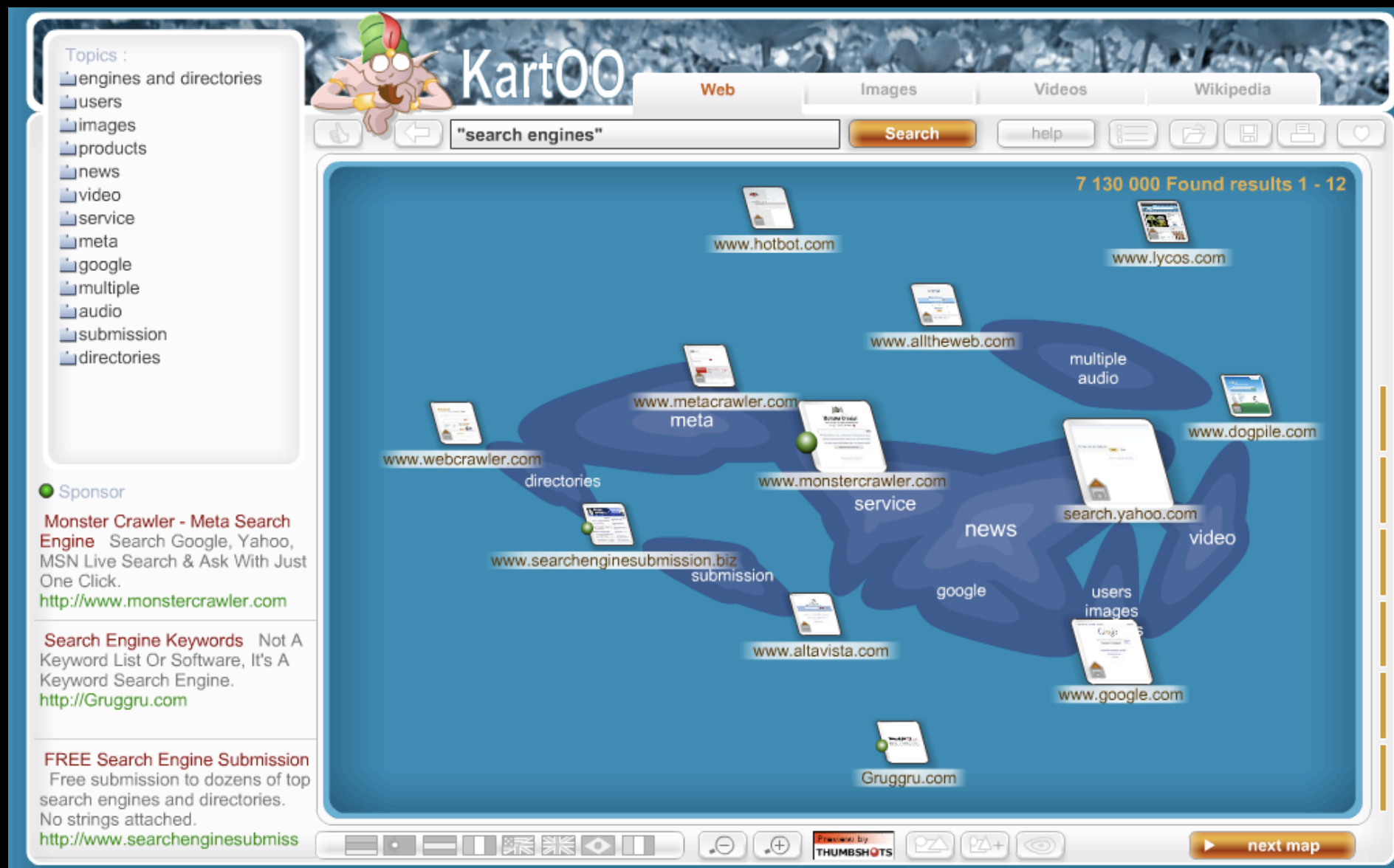
Dovrete specificare il vostro **scopo**, ma anche i **limiti** della vostra ricerca. Mi raccomando: state attenti quando la limiterete a specifici tipi di documenti.

Esempi di limitazioni con google: -intitle:, -allintitle:, -inurl:, -allinurl:, -allintext:, -inanchor;; con yahoo: -path:, -inurl:, etc.

Un esempio classico di ricerca “intelligente” che mira a sfronare la query:

"title:index title:of" -originurlextension:htm -originurlextension:html -papers -copyright +Oreilly

Per individuare più termini e concetti rilevanti ai fini della vostra ricerca, può essere utile usare motori di ricerca “grafici” come ad esempio: **Kart00** (<http://www.kartoo.com>), Touch, Ujiko, Dicy o Mooter).



3. Fate ricerche preliminari, per sondare il terreno...

Usate almeno 3 motori di ricerca principali:
<http://www.searchlores.org/main.htm>

Con quali cominciare?

GOOGLE,
FAST/ALLTHEWEB
e TEOMA

GOOGLE: www.google.com

lo conoscete tutti. Anche la **ricerca avanzata?**

Disolito lo usiamo mettendo una (!) parola. E' una strategia time-consuming e che espone a un eccesso di informazione...

(per saperne di più: <http://www.searchlores.org/google.htm>
http://www.google.com/advanced_search.html)

e googlescholar?

(<http://scholar.google.it/>)

FAST/ALLTHEWEB: <http://www.alltheweb.com>

(per saperne di più: <http://www.searchlores.org/fast.htm>)

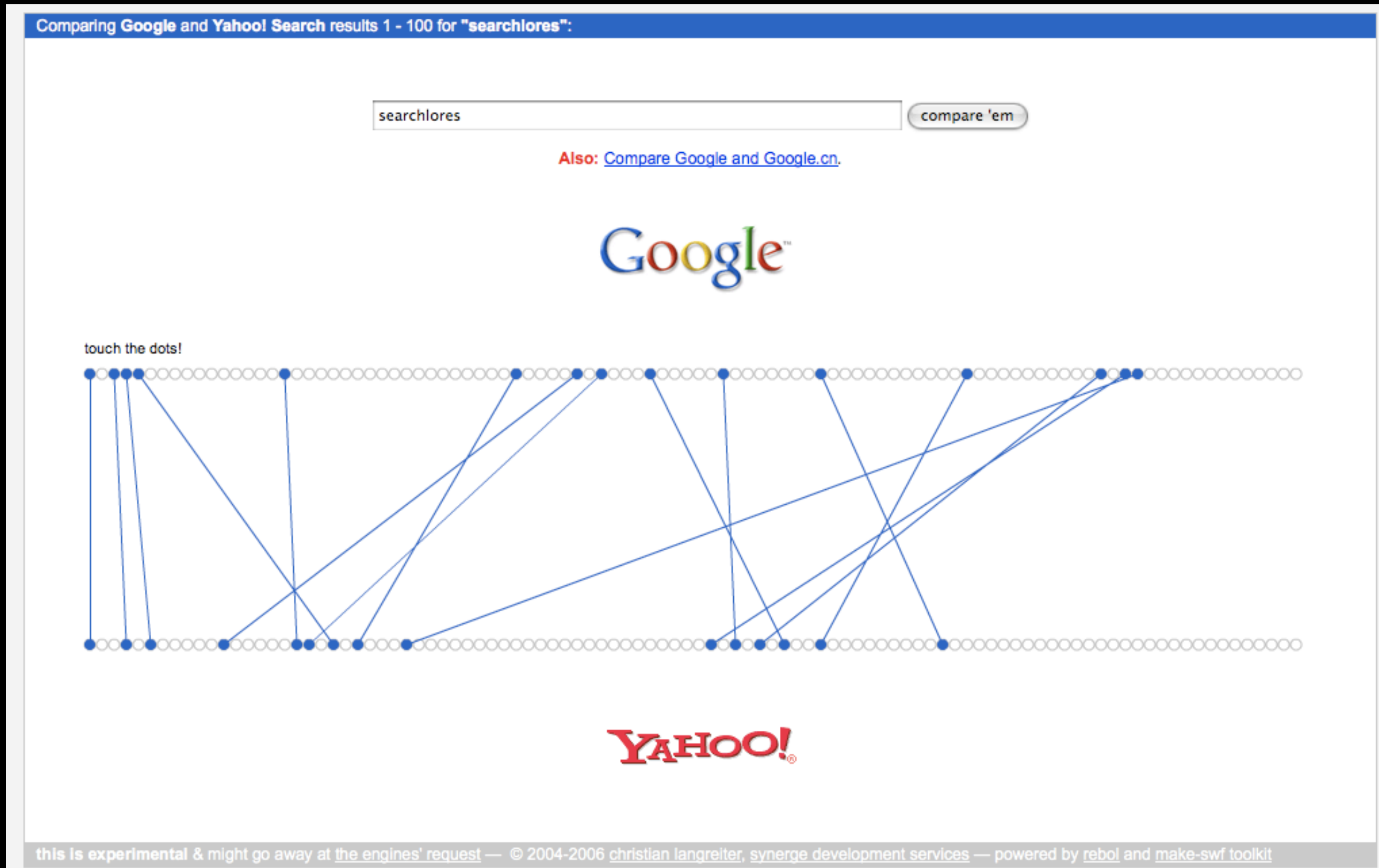
TEOMA: <http://it.ask.com/?o=312&l=dir>

(per saperne di più: <http://www.searchlores.org/teoma.htm>)

ALTAVISTA:

(per saperne di più: <http://www.searchlores.org/altavil.htm>)

Controllando i risultati, vi accorgete che non si sovrappongono più di tanto.



<http://www.langreiter.com/exec/yahoo-vs-google.html?q=searchlores>

Verificate se avete abbastanza concetti e query. Attenzione ai compromessi (è questione di esperienza).

Ampliando la vostra ricerca il numero dei risultati aumenterà.

Spesso è utile cominciare con *querystring* (collezioni di termini) specifiche e importanti e con quelle che occorrono meno. Non è necessario includere nella *query* tutti i concetti identificati nel vostro piano. Aumentando i termini nella *query*, il numero dei risultati diminuirà.

Saranno la vostra immaginazione nello scegliere o cambiare i termini delle *query* facendo ricerche incrociate, e la vostra capacità di adattare la scelta di termini al campo di dati che state cercando (tutto il web? uno db specifico? una biblioteca on-line?) a farvi trovare o meno quello che cercate.

NB: se cercate interi corpus di testi, metterete molti termini e sarete più specifici possibile; se cercate tra titoli (per esempio, il db di titoli di una rivista) dovrete stare molto più attenti quando aggiungete termini. Il numero di termini incluso nella query dovrebbe generalmente essere proporzionale alla quantità di dati che cercate.

In questa fase della ricerca, è come se tiraste le frecce un po' a caso, sperimentando diverse combinazioni terminologiche. In pratica, fate una ricerca, esaminate i risultati, studiate i più rilevanti e poi riformulate la ricerca per migliorarne l'efficacia.

NB: apportate un cambiamento alla volta, senno' rischiate di non capire quali sono i termini che hanno modificato i risultati (le cause) e di perdere i risultati delle ricerche precedenti.

4. Entriamo nel vivo della ricerca: esploriamo il web profondo

“combing”: setacciare il web cercando persone che hanno già fatto ricerche sul nostro argomento e individuare le “authorities”.

(per saperne di più: www.searchlores.org/combing.htm)

In una **ricerca scientifica**, ad esempio, cerchiamo:

- tutte le **riviste elettroniche** rilevanti per la nostra materia
- il maggior numero di **messaggi e newsgroup** che si occupano di argomenti affini al nostro
- tutti i **libri** che siamo in grado di rintracciare.

Troviamo queste fonti nelle **citazioni, bibliografie e link** all'interno delle fonti che troviamo (processo ricorsivo).

Articoli non accessibili gratuitamente possono essere su altri siti in giro per il web.

Importante: contattare esperti

Come nel mondo della stampa, anche sul web il successo delle ricerche riposa su una combinazione di esperienza e conoscenza del dominio. Per avere successo nel trovare le informazioni, bisogna essere (e/o diventare) esperti del campo di ricerca. Come? Per esempio diventando capaci di identificare, valutare e stimare gli esperti nel vostro campo attraverso messageboards, webrings, articoli, reensioni, libri.

Valutare è la parte più importante del processo di ricerca.

Un esempio: Dr Jeckill e Mr Hyde (Guédon). Email, report, interviste, articoli di giornale, immagini possono contribuire alla vostra ricerca più di libri “famosi” o “da citare per forza”.

5. Identificate
letteratura e fonti
rilevanti, inedite o poco
distribuite (cosiddette
“aree grigie”)

Questo è molto importante. Oggi gli articoli e i libri spesso sintetizzano risultati precedenti, già pubblicati o diffusi (condivisi) sul web.

Le aree grigie del web (articoli e proceeding di conferenze, tesi e dissertazioni non pubblicate, messaggi e comunicazioni non ufficiali, blog) molto spesso offrono informazioni di eccellente qualità.

Questa è una delle ragioni per cui potrebbe essere abbastanza pericoloso restringere la ricerca ai file doc o pdf: un'enorme quantità di informazioni è pubblicata in html.

Come realizzano tutti i ricercatori (presto o tardi) c'è molto da imparare da persone che sono parzialmente o totalmente estranee all'accademia. Anche per questo, ovviamente, i vostri strumenti di valutazione sono di importanza critica (in tutti i sensi).

In più ci sono "**web tides**": momenti in cui d'incontro tra accademici ed esperti "liberi" (vedi fenomeno del **web 2.0**). Esistono molte micro-comunità che si occupano delle stesse cose senza saperlo.

6. Rivedete il piano iniziale
regolarmente per
incorporare nuove
scoperte. Provando
diversi approcci, se
necessario

Cercare è un processo. E' di primaria importanza che aggiorniate costantemente la vostra struttura di riferimento, usando il materiale che avete trovato.

Dovreste buttar giù possibili approcci promettenti alternativi (fravia: for a rainy day ;-)).

Inoltre, per ricerche davvero “di lungo termine”, ricordate, quando rivedete i risultati, di fare una *checklist* delle fonti cercate e degli approcci che avete usato, per evitare di perdere tempo nel duplicare i tentativi.

Una buona idea è fare 2 volte la vostra ricerca senza usare mai google. Un'altra possibilità è provare a usare approcci differenti, tipo ricerche su ftp, usenet & blog, irc.

7. Rifate le vostre query in diverse lingue

Qui un elenco di motori di ricerca locali e regionali:
<http://www.searchlores.org/regional.htm>

Qui una pagina sulla ricerca multilingue:
<http://www.searchlores.org/eurosearch.htm>

NB. Conoscere più di una lingua per chi fa ricerca sul web è quasi una conditio sine qua non (beh, non solo sul web;-))

8. Come sapere se la vostra ricerca è stata efficace?

L'efficacia di una data ricerca si misura in termini di precisione (quanti dei documenti trovati sono rilevanti per la vostra query? 100%? 50%? 10%?) e di richiamo (quante di tutte le fonti teoricamente disponibili sono state trovate attraverso la vostra specifica query? 100%? 50%? 10%?). Sul web però queste domande, la seconda soprattutto, non hanno senso.

Un consiglio utile: chiedere ai propri amici o collaboratori di fare le stesse ricerche. Ogni essere umano segue percorsi e piste differenti per ricercare nello stesso insieme di dati. Anche solo consultarvi con altri può aiutarvi a migliorare il risultato generale.

esempi di errori...

- Errori ortografici
- errori terminologici e concettuali
- errori nell'uso degli operatori booleani (AND e OR)
- dimenticare risorse importanti
- errori di valutazione
- insufficiente regionalizzazione della query
-

Ricordate che spesso aver dimenticato o aver sbagliato qualcosa può essere una fortuna :-)

Programma del corso

Lezione 1. I motori di ricerca

- (1). La topologia del web
- (2). In pratica: Strumenti, regole, consigli
 - a. Ricerche di breve durata

Lezione 2. I motori di ricerca

- (2). In pratica: Strumenti, regole, consigli
 - b. Ricerche scientifiche (di lunga durata)

Lezione 3. Strumenti per selezionare e archiviare i risultati

- (1). Web semantico e web 2.0
- (2). Strumenti collaborativi per la ricerca

Lezione 4. Open Access

- (1). Storia e filosofia
- (2). Strumenti, problemi, prospettive

il web si basa su 3 principi

URI/URL

HTTP

HTML

L'evoluzione del Web: 1990-oggi

Partito con una pagina ai primi anni 90, nel 2005 il WWW contava più di 100 milioni di siti, e 8 miliardi di pagine pubblicamente accessibili.

Perché?

- ➔ Non ci sono limiti al numero di link e al numero di pagine che si possono aggiungere.
- ➔ Per farlo, non bisogna coordinarsi con nessuna autorità centrale.

Il web ha dei limiti architettonici che fanno sì che le macchine non siano in grado di interpretare molti dei dati che sono presenti.

Perché?

2007

Il web è fatto di

siti, pagine e **documenti**

noi (umani) lo vediamo
così



Bollettino telematico di filosofia politica:
ProgettoMaster

ProgettoMaster

Cerca

Titoli

Testo

ModificheRecenti

CercaPagina

AiutoContenuti

ProgettoMaster

Modifica (modo testuale)

Modifica (modo grafico)

Commenti

Informazioni

Allegati

Altre azioni:

Le cinque fasi della produzione scientifica (5 ore di didattica frontale - Maria Chiara Pievatolo)

- Registrazione (modalità di pubblicazione)

1. Dalle comunità di conoscenza alla stampa
2. Dalla stampa alla rete

- Certificazione (valutazione della ricerca)

1. Origine e fortuna del peer review
2. Il fattore d'impatto e altri indici bibliometrici
3. Le alternative possibili: il soft peer review in un ambiente ad accesso aperto

- Diffusione

1. I vincoli tecnologici ed economici
2. I monopoli intellettuali

- Archiviazione

- ### 1. Biblioteche istituzionali, biblioteche diffuse

- Riconoscimento

1. Ai posteri l'ardua sentenza: come conciliare certificazione e riconoscimento?

le macchine lo vedono così

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8">
<meta name="robots" content="index,nofollow">

<title>ProgettoMaster - Wiki del Bollettino telematico di filosofia politica</title>
<script type="text/javascript" src="/moin_static161/common/js/common.js"></script>

<script type="text/javascript">
<!--
var search_hint = "Cerca";
//-->
</script>

<script type="text/javascript">
<!-- // GUI edit link and il8n
var gui_editor_link_href = "/moinwiki/ProgettoMaster?action=edit&editor=gui";
var gui_editor_link_text = "Modifica (modo grafico)";
//-->
</script>

<link rel="stylesheet" type="text/css" charset="utf-8" media="all" href="/moin_static161/modern/css/common.css">
<link rel="stylesheet" type="text/css" charset="utf-8" media="screen" href="/moin_static161/modern/css/screen.css">
<link rel="stylesheet" type="text/css" charset="utf-8" media="print" href="/moin_static161/modern/css/print.css">
<link rel="stylesheet" type="text/css" charset="utf-8" media="projection" href="/moin_static161/modern/css/projection.css">

<!-- css only for MSIE browsers -->
<!--[if IE]>
    <link rel="stylesheet" type="text/css" charset="utf-8" media="all" href="/moin_static161/modern/css/msie.css">
<![endif]-->

<link rel="Start" href="/moinwiki/PaginaPrincipale">
<link rel="Alternate" title="Formattazione Wiki" href="/moinwiki/ProgettoMaster?action=raw">
<link rel="Alternate" media="print" title="Versione stampabile" href="/moinwiki/ProgettoMaster?action=print">
<link rel="Search" href="/moinwiki/CercaPagina">
<link rel="Index" href="/moinwiki/IndiceDeiTitoli">
<link rel="Glossary" href="/moinwiki/IndicePerParola">
<link rel="Help" href="/moinwiki/AiutoSuFormattazione">
</head>

<body lang="it" dir="ltr">

<div id="header">
<div id="logo"><a href="/moinwiki/PaginaPrincipale">
</div>
<div id="searchform"><form id="searchform" method="get" action="/moinwiki/ProgettoMaster">
<div>
```

Il Web Semantico:

Premessa - i limiti di HTML

Un documento HTML è un file di testo che contiene: Testo + Markup

- Confonde l'aspetto di un documento e la semantica del testo
- Comprende la possibilità di overlapping

HTML

non permette

lo scambio di informazione strutturata
tra le macchine

Cioè....

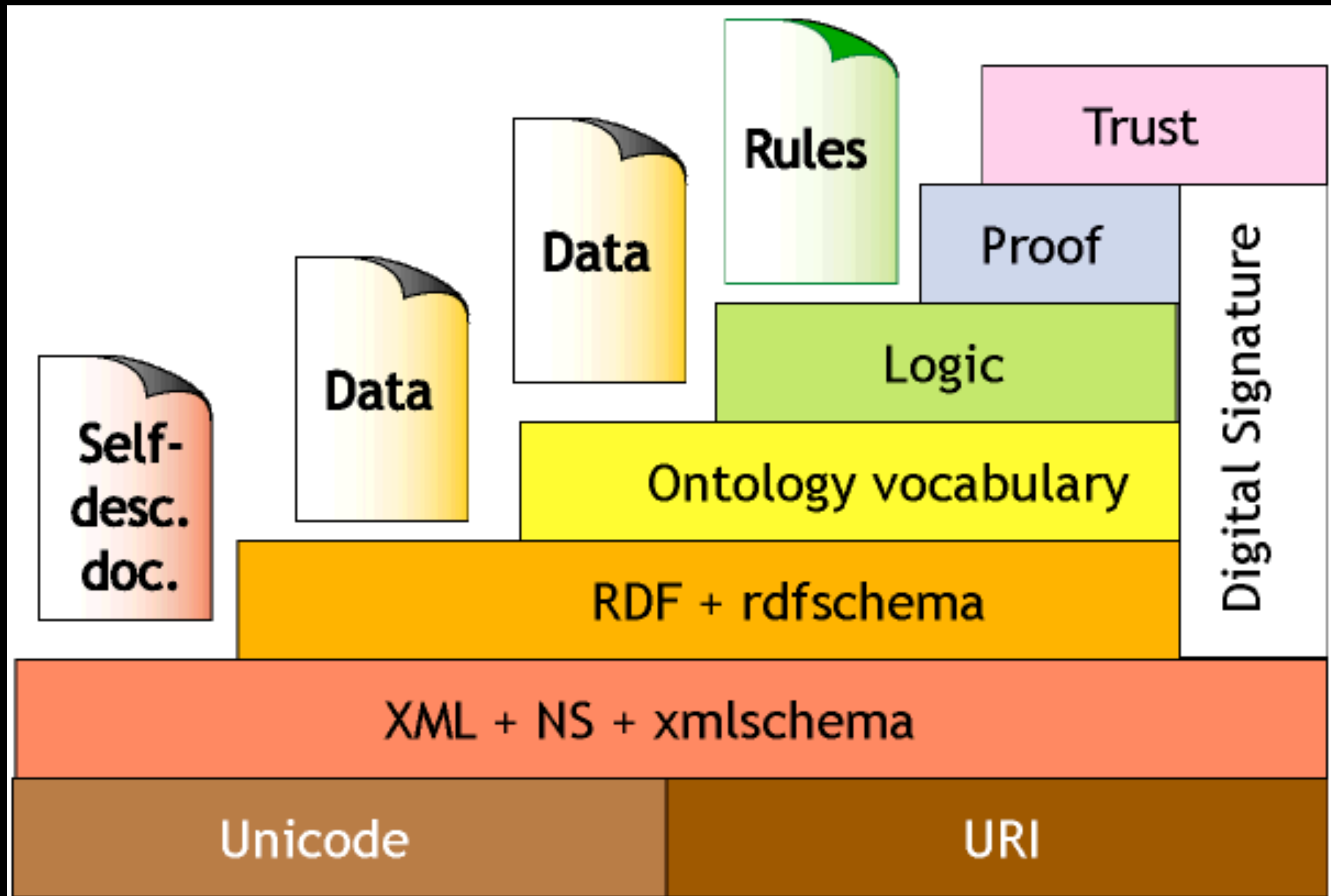
- Si pubblicano documenti
- pensati per essere utilizzati dalle persone, non dalle macchine
- Uno spazio dell'informazione segmentato e disconnesso
- Il significato dei documenti e dei link è implicito e interpretabile solo dalle persone

Lo spazio dell'informazione del Web è stato progettato con l'obiettivo che fosse utile non solo per la comunicazione tra umani, ma perché vi partecipassero anche le macchine. Il “Web Semantico”, al fine di rendere i dati comprensibili da appositi software, **accentua la separazione tra “data”** (utilizzabili ed elaborabili dalle macchine) e **“documents”** (leggibili dagli umani); per farlo, il W3C ha dato vita a **nuovi formati di codifica (XML, RDF)**, tali da poter esprimere le informazioni in forma processabile dalla macchina.

<<http://bfp.sp.unipi.it/~didonato/ovre/telema/ws.html>>

II Web Semantico:

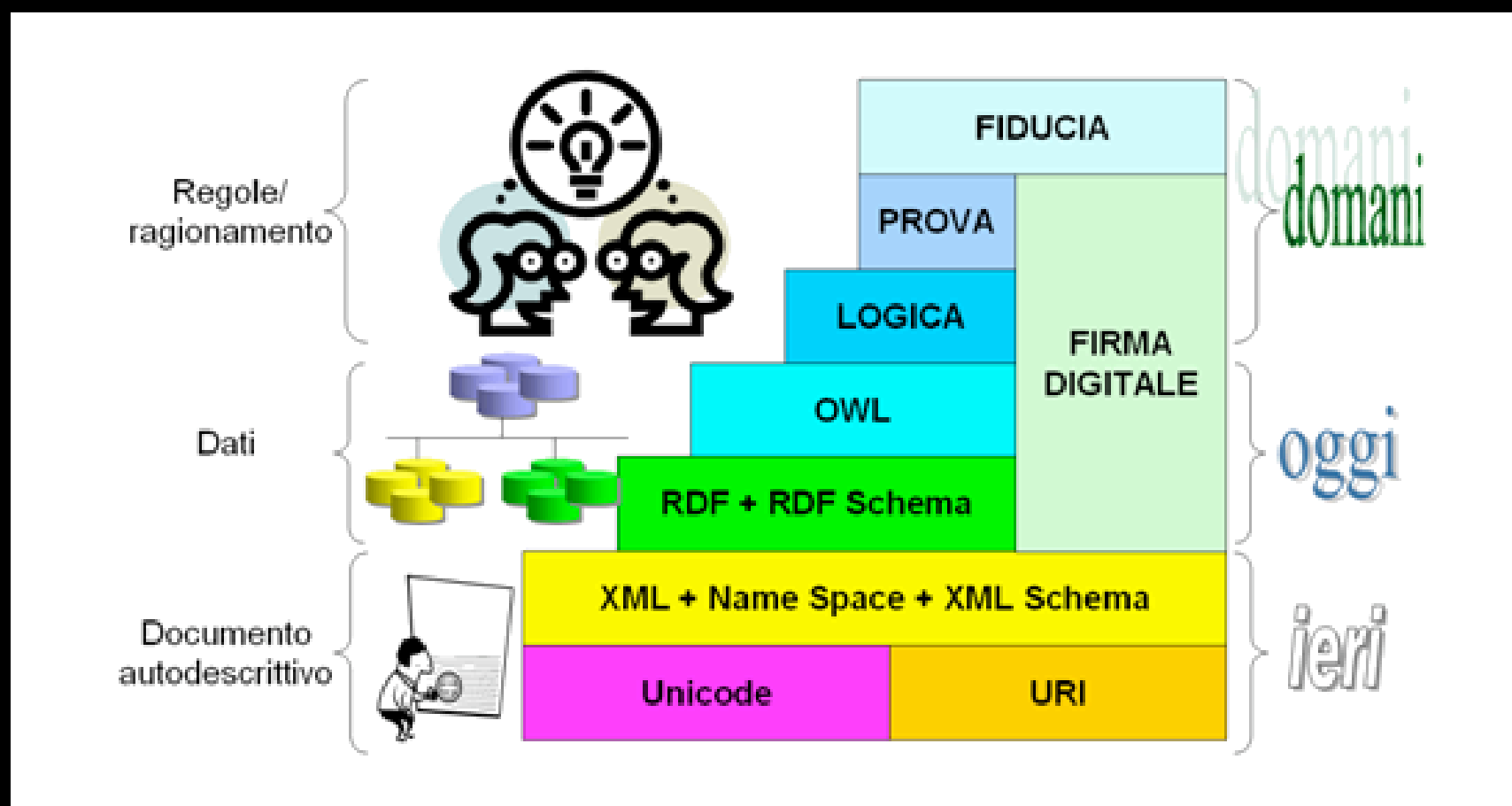
architettura - “layer cake”



(meta)dati

I metadati, “dati sui dati”, sono informazioni che arricchiscono le pagine Web, così che software appositamente creati possano farne uso. Nuove informazioni, strutturate tramite metadati, esprimono il contesto dei dati stessi – sono, cioè, informazioni aggiuntive sull’informazione. Sono metadati, ad esempio, l’autore e il titolo di un articolo.

Schema Web Semantico



<http://www2.autistici.org/bakunin//doc/wsws/mgp00012.html>

Per saperne di più:

<http://bfp.sp.unipi.it/~didonato/ovre/cagliarifad2.pdf>

L'evoluzione sociale del World Wide Web: il Web 2.0

“Web 2.0 is a set of economic, social, and technology trends that collectively form the basis for the next generation of the Internet - a more mature, distinctive medium characterized by user participation, openness, and network effects”.

Tim O'Reilly, 2004.

Web 2.0: una definizione, sette caratteristiche

1) "Il web come piattaforma"

Un esempio: Google account <http://www.google.it>

2) Il software è un servizio e non un prodotto

3) Decentralizzazione: ogni client è anche un server (P2P)

Esempi: Napster, Emule, etc..

4) Centralità dei dati

User Generated Contents

I contenuti sono generati dagli utenti e condivisi all'interno della comunità.

In questo modo, sono gli stessi utenti a creare il “servizio”.

Qualche esempio: Youtube, Flickr, ...

Facciamo una prova...

<http://www.flickr.com/search/?q=pisa+assemblea+universit%C3%A0>

1 gennaio 2007



Cioè?



CASI UMANI
#1:
L'ENTUSIASTA
VS
LO SCETTICO

SINDROME WEB.2.0

CASI UMANI 2.0 PRESENTATI
DALL'ESIMO DR. TRAM



5) Intelligenza collettiva

*** si creano reti sociali** tra persone che condividono gli stessi interessi.

Esempi: My Space, Facebook

*** si diffondono "folksonomie" (vs. tassonomie):**
stili di categorizzazione collaborativa che utilizzano parole chiave liberamente scelte, spesso definite tag.



Un esempio:
del.icio.us
<http://del.icio.us>

* Wikipedia - un esempio particolare di intelligenza collettiva

Web 2.0 - Wikipedia, the free encyclopedia

http://en.wikipedia.org/wiki/Web2.0

Come iniziare | Ultime notizie | Procedura colf e bad... | DENUNCIA DI RAPPO... | TuttoInps - I lavorat...


Recently Bookmarked | YouTube - Non è l'a... | L'editoria delle Unive... | Academic Proofreadi... | ivy on concrete and ... | red ivy #2 | Poison Ivy's House | Ivy | Ivy >>

Disable | Cookies | CSS | Forms | Images | Information | Miscellaneous | Outline | Resize | Tools | View Source | Options

Web 2.0 - Wikipedia, the free ...

Sign in / create account

article | discussion | edit this page | history

 **WIKIPEDIA**
The Free Encyclopedia

navigation

- Main page
- Contents
- Featured content
- Current events
- Random article

interaction

- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

search

Go Search

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link
- Cite this article

 **You** can help Wikipedia change the world!

32,833 have donated.

[» Donate now!](#)

From the fundraising blog – [Support \(Or Join\) Wikimedia's New Team in San Francisco](#)

"Wikipedia makes me more productive!" – Harald Lie

Web 2.0

From Wikipedia, the free encyclopedia
(Redirected from [Web2.0](#))


Web 2.0 refers to a perceived second [generation](#) of web-based communities and [hosted services](#) — such as [social-networking sites](#), [wikis](#), and [folksonomies](#) — which aim to facilitate [creativity](#), collaboration, and sharing between users. The term gained currency following the first [O'Reilly Media Web 2.0 conference](#) in 2004.^{[2][3]} Although the term suggests a new version of the [World Wide Web](#), it does not refer to an update to any technical specifications, but to changes in the ways [software developers](#) and [end-users](#) use webs. According to [Tim O'Reilly](#),

"Web 2.0 is the [business](#) revolution in the [computer industry](#) caused by the move to the [Internet](#) as [platform](#), and an attempt to understand the rules for success on that new platform."^[4]

Some technology experts, notably [Tim Berners-Lee](#), have questioned whether one can use the term in a meaningful way, since many of the technology components of "Web 2.0" have existed since the early days of the Web.^{[5][6]}

Contents [\[hide\]](#)

- Defining "Web 2.0"
- Characteristics
- Technology overview
- Innovations sometimes associated with "Web 2.0"
 - Web-based applications and desktops
 - Rich Internet applications
 - XML and RSS

 **Web 2.0**

On September 30, 2005, [Tim O'Reilly](#) wrote a [piece](#) summarizing his view of Web 2.0. The [mind-map](#) pictured above (constructed by [Markus Angermeier](#) ^[1] on November 11, 2005) sums up some of the [memes](#) of Web 2.0, with example-sites and services attached.

6) Some rights reserved (copyleft)

Una gestione attenta del copyright permette il riuso dei contenuti: la diffusione di licenze creative commons facilita la circolazione la diffusione di molti contenuti digitali.



Ancora un esempio: ricerca avanzata su Flickr

<http://www.flickr.com/search/advanced/?q=pisa%20assemblea%20universit%C3%A0>



worldwide

Seleziona un sito

Cerca

Vai

Creative Commons, un copyright flessibile per opere creative

Cerca e Pubblica

🎵 **Audio**
Musica, suoni, discorsi...

📷 **Immagini**
Fotografie, immagini, illustrazioni...

🎬 **Video**
Film, animazioni...

📖 **Testi**
Libri, blog, saggi...

🎓 **Education**
Lezioni, corsi, libri di testo...



Cerca
musica, fotografie, e altro



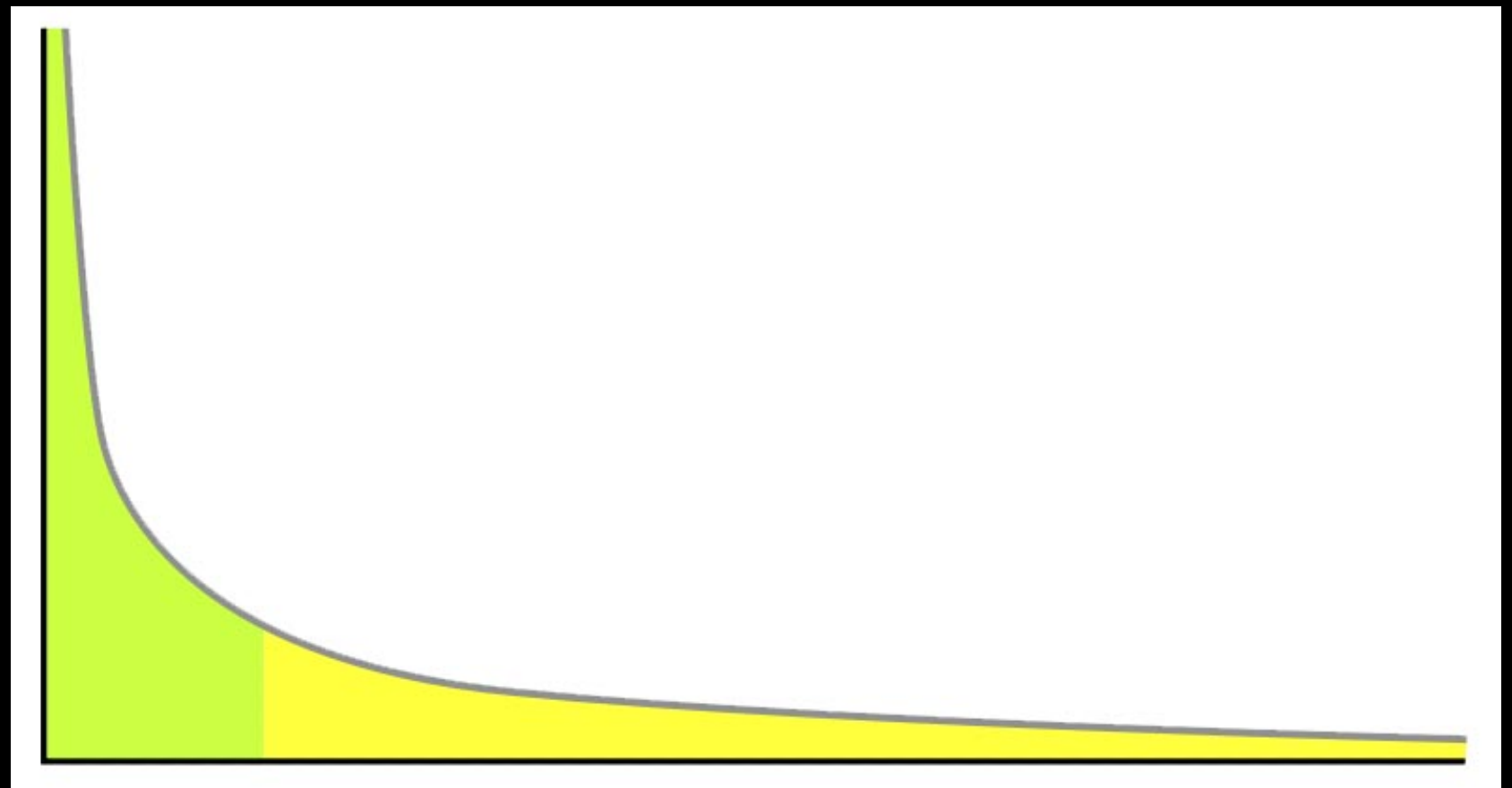
Pubblica
le tue creazioni, legalmente

Le licenze Creative Commons offrono sei diverse modulazioni dei diritti previsti dalla legge sul diritto d'autore per artisti, giornalisti, docenti, istituzioni, e creatori in genere che desiderino condividere in maniera ampia

7) La “coda lunga” (long tail)

presi complessivamente i (tanti) mercati nicchia
possono essere più remunerativi

dei (pochi)
grandi
business.



[The Long Tail by Chris Anderson]

Esempi di siti Web2.0

LibraryThing

<http://www.librarything.com/>

FutureMe

<http://www.futureme.org/>

Zopa

<http://www.zopa.it/ZopaWeb/>

delicious

<http://delicious.com//>

Connotea

<http://www.connotea.org/>

CiteULike

<http://citeulike.org/>

delicious

<http://delicious.com/>

citeulike

<http://citeulike.org/>

citeulike 

sponsored by  Springer

Logged in as [gioia](#) | [Log Out](#) | [FAQ](#) | [\[?\]](#)

Search

CiteULike ▾

Post Article ▾

My CiteULike ▾

Journals ▾

Groups ▾



www.academia.edu
Sign Up!



Everyone's library

Some recent papers posted to CiteULike - all mixed together. You can also see [just your library](#).

- [Medical predictors for psychopathology in adults with operated congenital heart disease](#)
Eur Heart J, Vol. 25, No. 18. (2 September 2004), pp. 1605-1613.
by EHM [van Rijen](#), EMWJ [Utens](#), JW [Roos-Hesselink](#), FJ [Meijboom](#), RT [van Domburg](#), JRTC [Roelandt](#), AJJC [Bogers](#), FC [Verhulst](#)
posted to [no-tag](#) by [tzellofouska](#) on 2008-10-16 15:41:10 as ★★
- [A revised simplex search procedure for stochastic simulation response-surface optimization](#)
(1998), pp. 751-760.
by David G [Humphrey](#), James R [Wilson](#)
posted to [uncertainty optimization](#) by [vbaudoui](#) on 2008-10-16 15:40:32 as ★★
- [P2P Experimentations with Splay: From Idea to Deployment Results in 30 min.](#)
Peer-to-Peer Computing, IEEE International Conference on, Vol. 0 (2008), pp. 189-190.
by Lorenzo [Leonini](#), ?tienne [Rivi?re](#), Pascal [Felber](#)
posted to [splay p2p evaluation deployment](#) by [ansobe](#) on 2008-10-16 15:40:22 as ✓
- [Longitudinal development of psychopathology in an adult congenital heart disease cohort](#)
International Journal of Cardiology, Vol. 99, No. 2. (18 March 2005), pp. 315-323.
by E [Vanrijen](#), E [Utens](#), J [Rooshesselink](#), F [Meijboom](#), R [Vandomburg](#), J [Roelandt](#), A [Bogers](#), F [Verhulst](#)
posted to [no-tag](#) by [tzellofouska](#) on 2008-10-16 15:40:21 as ★★
- [It's time to pay attention to attention in aging.](#)
Learning & memory (Cold Spring Harbor, N.Y.), Vol. 9, No. 4. (g 2002), pp. 151-152.
by J [McGaughy](#), H [Eichenbaum](#)
posted to [rat idseds human gk decline in performance commentary](#) by [cvq](#) on 2008-10-16 15:39:11 as ★★



Everyone's Tags

Most active tags on CiteULike

Filter:

2008 algorithm analysis
and annotation architecture
article bayesian
bioinformatics biology
brain c_elegans
caenorhabditis_elega
cancer celegans cell
citeulike classification
clustering cognition
community complexity
culture data database
design development dna
dynamics economics
education elegans
evaluation evolution
expression fmri
gazette_abstract gene
genetics genome
genomics health history
human imaging in
inflammation information

gestire e condividere bibliografie

CiteULike: My library [178 articles]

http://citeulike.org/user/gioia

Logged in as [gioia](#) | [Log Out](#) | [FAQ](#) | [\[?\]](#)

citeulike | sponsored by Springer

[CiteULike](#) ▾ [Post Article](#) ▾ [My CiteULike](#) ▾ [Journals](#) ▾ [Groups](#) ▾

[Library](#) [Groups](#) [Watchlist](#) [Profile](#) [Blog](#)

[Recent](#) | [Unread](#) | [Search](#) | [Authors](#) | [Tags](#) | [Export](#) | [Import](#)

My library [178 articles]

Recent papers added to My library. [watch](#)

- [Evaluation of Digital Library Impact and User Communities by Analysis of Usage Patterns](#)
D-Lib (2000)
by J Bollen
posted to [no-tag](#) by [gioia](#) on 2008-03-10 14:46:01 as ★★ [along with 1 person](#)
- [Networks of Trust: Scholarly Correspondence and Scientific Exchange in Early Modern Europe](#)
The Medieval History Journal, Vol. 6, No. 1. (1 April 2003), pp. 1-32.
by Franz Mauelshagen
posted to [republic_of_science](#) by [gioia](#) on 2008-02-25 11:28:00 as ★★
- [Charisma, Recognition, and the Motivation of Scientists](#)
The American Journal of Sociology, Vol. 78, No. 5. (1973), pp. 1119-1134.
by Bernard H Gustin
posted to [republic_of_science](#) by [gioia](#) on 2008-02-22 15:31:09 as ★★

gioia's tags
All tags in gioia's library

Filter:

academies alp architectonic
authorship children
citation_indicators
computability computer_science
condorcet copyright credit
digital_libraries
e-journals england
enlightenment fichte
freedom_of_press
gelehrtenrepublik
germany
history_of_philosophy
history_of_press
history_of_science

Springer eBook Collection

CHECK OUT ACADEMIA.EDU, IT'S LIKE AN ACADEMIC FACEBOOK

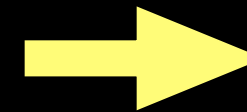
www.academia.edu
Sign Up!

ANNUAL REVIEWS
INSIGHTFUL RESEARCH STARTS HERE

citeulike

Per cominciare a usare
citeulike è sufficiente
registrarsi cliccando su
register
presente in ogni pagina

Tag cloud



gioia's tags

All tags in gioia's library

Filter:

academies architectonic
citation_indicators
computability computer_science
condorcet copyright credit
e-journals england
enlightenment fichte
freedom_of_press
gelehrtenrepublik
germany
history_of_philosophy
history_of_press
history_of_science
history_of_university
humboldt hyperjournal
hypertext internet ip journals
kant kant_bibliography
klopstock krv
learned_journals
libraries open_access
peer_review
perpetual_peace philosophy
political_philosophy
prussia
public_use_of_reason
reimarus
republic_of_letters
republic_of_science sci science
scientific_publishing

citeulike

[Register](#) | [Log in](#) | [FAQ](#) | | | | | | | | |

Article title

Search

Navigation

- [Home](#)
- [Log in](#)
- [Register](#)
- [Discussion list](#)

Journals

- [Browse current issues](#)

Groups

- [View group](#)

Experimental Features

- [Import from BibTeX](#)

Register for CiteULike

All you need to do is fill in the boxes below. It's really very painless.

Username:

Note: You can't have spaces or punctuation in your username

Password:

Confirm Password:

Note: Both your username and password are case sensitive.

Email:

Note: We will not spam you. We will not sell your email address to anyone. We will not give your email address to anyone else (including other users on the system) without your explicit consent. We hate spam as much as you.

Sign up

connotea

<http://www.connotea.org/>

Per concludere...
il Web 2.0 in un video

The Machine is Us/ing Us
<<http://www.youtube.com/watch?v=6gmP4nk0EOE>>

Biblio-linkografia

Times, copertina del 3 gennaio 1983

<http://www.time.com/time/covers/0,16641,19830103,00.html>

Times, copertina del primo gennaio 2007

<http://www.time.com/time/covers/0,16641,20061225,00.html>

T. Berners-Lee, "The World Wide Web - Past, present, future. Exploring Universality" (2002)

<http://www.w3.org/2002/04/Japan/Lecture.html>

T. Berners-Lee, Hearing on the "Digital Future of the United States: Part I -- The Future of the World Wide Web", 1 marzo 2007

<http://dig.csail.mit.edu/2007/03/01-ushouse-future-of-the-web.html>

A. Marchesini, Workshop Web semantico (2005)

<http://www2.autistici.org/bakunin/doc/ws/ws/>

Web semantico: breve linkografia ragionata (2005-2007)

<http://bfp.sp.unipi.it/~didonato/ovre/telema/ws.html>

Tim O'Reilly, What Is Web 2.0, Design Patterns and Business Models for the Next Generation of Software

<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html?page=1>

(tr. it.) Cosa è Web 2.0. Design Patterns e Modelli di Business per la Prossima Generazione di Software

<http://www.xyz.reply.it/web20/>

Web 2.0 (voce di Wikipedia)

http://en.wikipedia.org/wiki/Web_2.0

F. Meschini, eContent: tradizionale, semantico o 2.0?, febbraio 2007

<http://dspace.unitus.it/handle/2067/162>

F. Meschini, Introduzione al [Web|Library] 2.0, Seminario autunnale Ciber 2007 20/11/2007 - Caspur

<https://cab.unime.it/ciber/?p=104>

M. Barbera, Semantic Web e Open Source per i beni culturali, LuBeC (Lucca) 2007, novembre 2007

Pausa...



Programma del corso

Lezione 1. I motori di ricerca

- (1). La topologia del web
- (2). In pratica: Strumenti, regole, consigli
 - a. Ricerche di breve durata

Lezione 2. I motori di ricerca

- (2). In pratica: Strumenti, regole, consigli
 - b. Ricerche scientifiche (di lunga durata)

Lezione 3. Strumenti per selezionare e archiviare i risultati

- (1). Web semantico e web 2.0
- (2). Strumenti collaborativi per la ricerca

Lezione 4. Open Access

- (1). Storia e filosofia
- (2). Strumenti, problemi, prospettive

0. Indice

1. Definizioni introduttive

2. Storia

3. Filosofia

4. Strumenti

5. Problemi e prospettive

6. Link e bibliografia

I. Definizioni introduttive/ I

“La letteratura ad accesso aperto (Open Access, OA) è digitale, online, gratuita e libera da buona parte delle restrizioni dettate dalle licenze per i diritti di sfruttamento commerciale. Queste condizioni sono possibili grazie ad Internet e al consenso dell'autore o del titolare dei diritti d'autore.”

Ci si riferisce alla **letteratura di ricerca**.

“L'accesso aperto è compatibile con il processo di referaggio (peer review).”

(P. Suber, Breve introduzione all'open access)

I. Definizioni introduttive/2

“La letteratura ad accesso aperto ha dei **costi di produzione**, sebbene inferiori a quelli delle pubblicazioni commerciali. La questione non riguarda come abbassarli, ma se vi siano dei modi di coprirli **che non ricadano sui lettori e non creino barriere all'accesso**.

I modelli economici dell'accesso aperto dipendono dalle modalità con cui lo si fornisce.”

(P. Suber, Breve introduzione all'open access)

I. Definizioni introduttive/3



“* Gli **archivi aperti o repositories** non effettuano referaggio ma rendono semplicemente i loro contenuti liberamente disponibili al mondo, siano essi **preprints** non referati o **postprints** referati, o entrambi. Gli archivi possono essere **istituzionali**, come quelli di università e centri di ricerca, o **disciplinari**, come quelli per la fisica o l'economia. Gli autori possono archiviare i propri preprints senza alcuna necessità di autorizzazione altrui, ed una gran parte delle riviste consente agli autori anche il deposito dei postprints. ...”

“ * Le **riviste ad accesso aperto** **effettuano il referaggio** e rendono i loro articoli selezionati liberamente disponibili al mondo. I loro costi consistono nella gestione dei processi editoriali, dal referaggio alla manipolazione dei manoscritti, allo spazio sul server. Le riviste ad accesso aperto coprono i propri costi in un modo molto simile alle radio e alle televisioni: coloro che hanno interesse alla disseminazione dei contenuti pagano anticipatamente i costi di produzione in modo che l'accesso alla lettura sia libero. ...”

(P. Suber, *Breve introduzione all'open access*)

Indice

1. Definizioni introduttive

2. Storia

3. Filosofia

4. Strumenti

5. Problemi e prospettive

6. Link e bibliografia

2. Storia > il movimento “open access” (1990-2005)

Esperienze

- P. Ginsparg: (xxx) ArXiv.org Los Alamos, 1991
- PubMed, PLOS

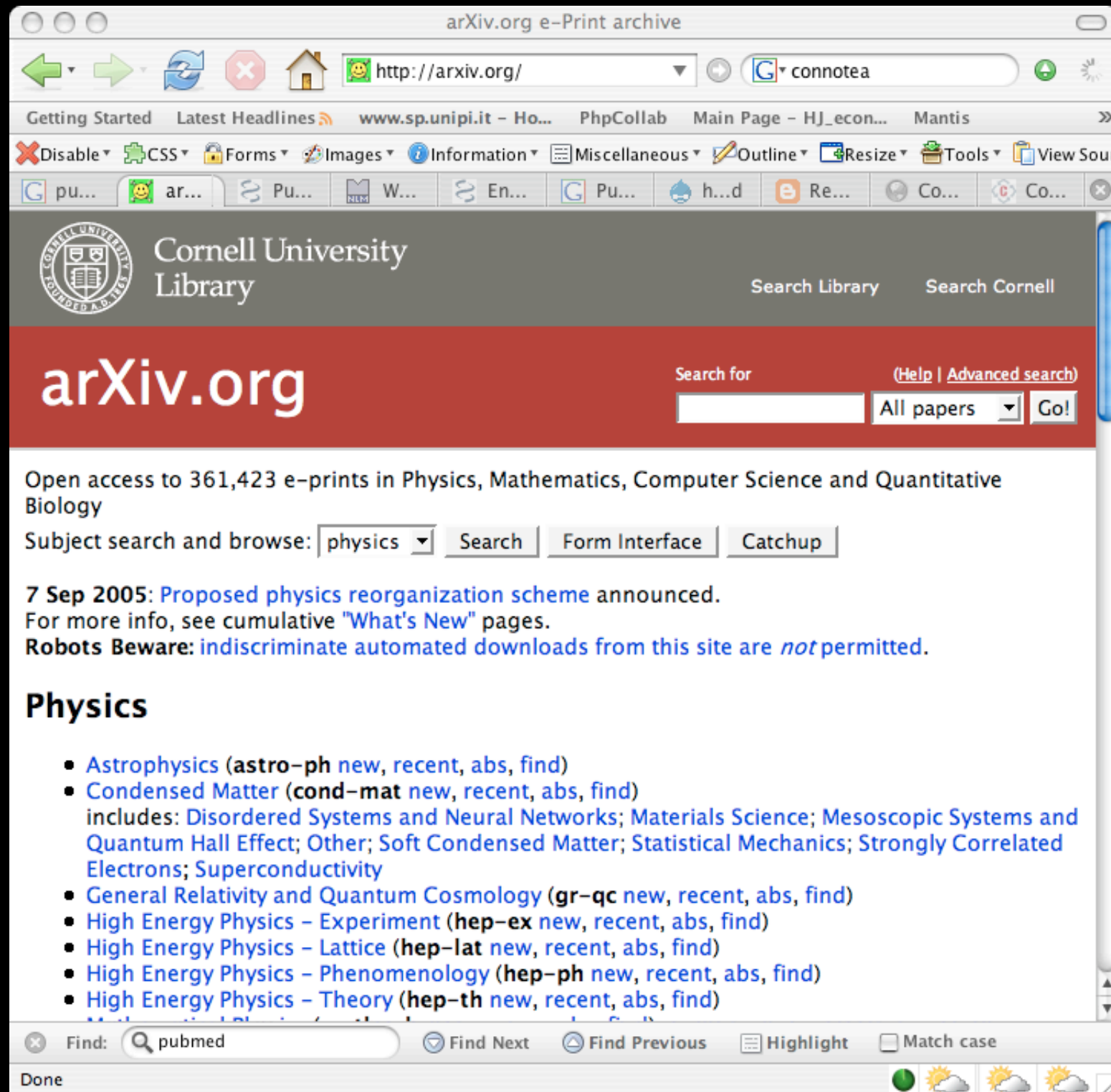
Tecnologia

- Open Archives Initiative, Santa Fe, 1999
- OAI-PMH v. 1.0, 2001 e v. 2.0, 2002

Politica

- Budapest Open Access Initiative, 2002
- Bethesda Statement on Open Access Publishing 2003
- Accesso aperto alla letteratura scientifica (Dichiarazione di Berlino) 2003
- Berlin 3: Progress in Implementing the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities 2005
- Berlin 4: From Promise to Practice (29-31 marzo 2006: OGGI!)

2. Storia > ArXiv.org

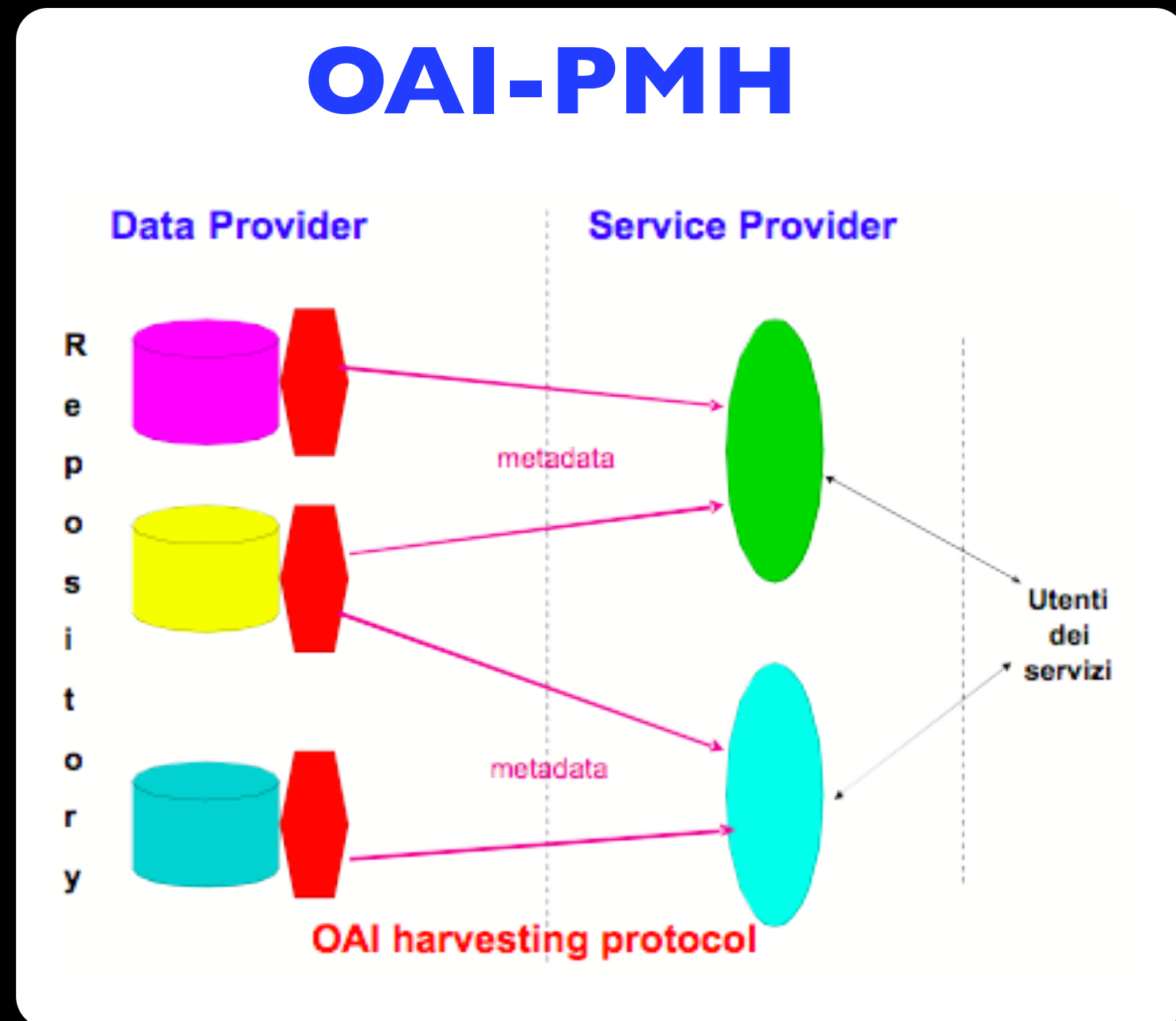


Paul Ginsparg

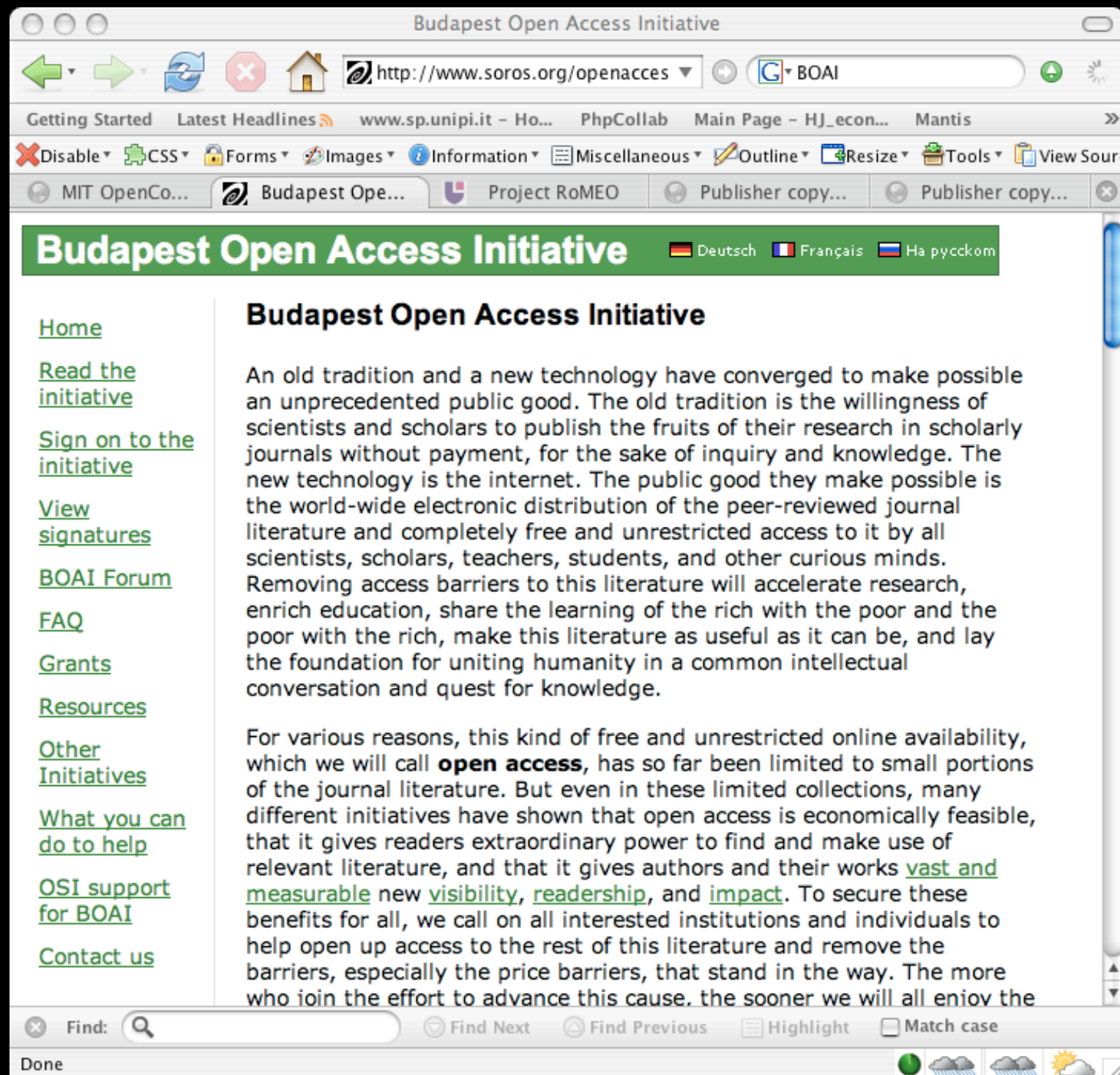
“Il bisogno di scambiare risultati scientifici è un fine diverso e complementare al modo in cui funziona il sistema delle pubblicazioni scientifiche”

2. Storia > Open Archives Initiative

- Sviluppo e promozione di **standard** per l'**interoperabilità**, per facilitare la disseminazione efficiente dei contenuti: incrementarne la disponibilità e facilitare il processo di comunicazione scientifica.
- L'**architettura** tecnologica fondamentale e gli **standard OAI** sono **indipendenti** sia dai **contenuti offerti** sia dal **meccanismo economico** che ruota attorno ai contenuti.
- **Dublin Core** è lo standard dei metadati



2. Storia > le istituzioni per l'open access



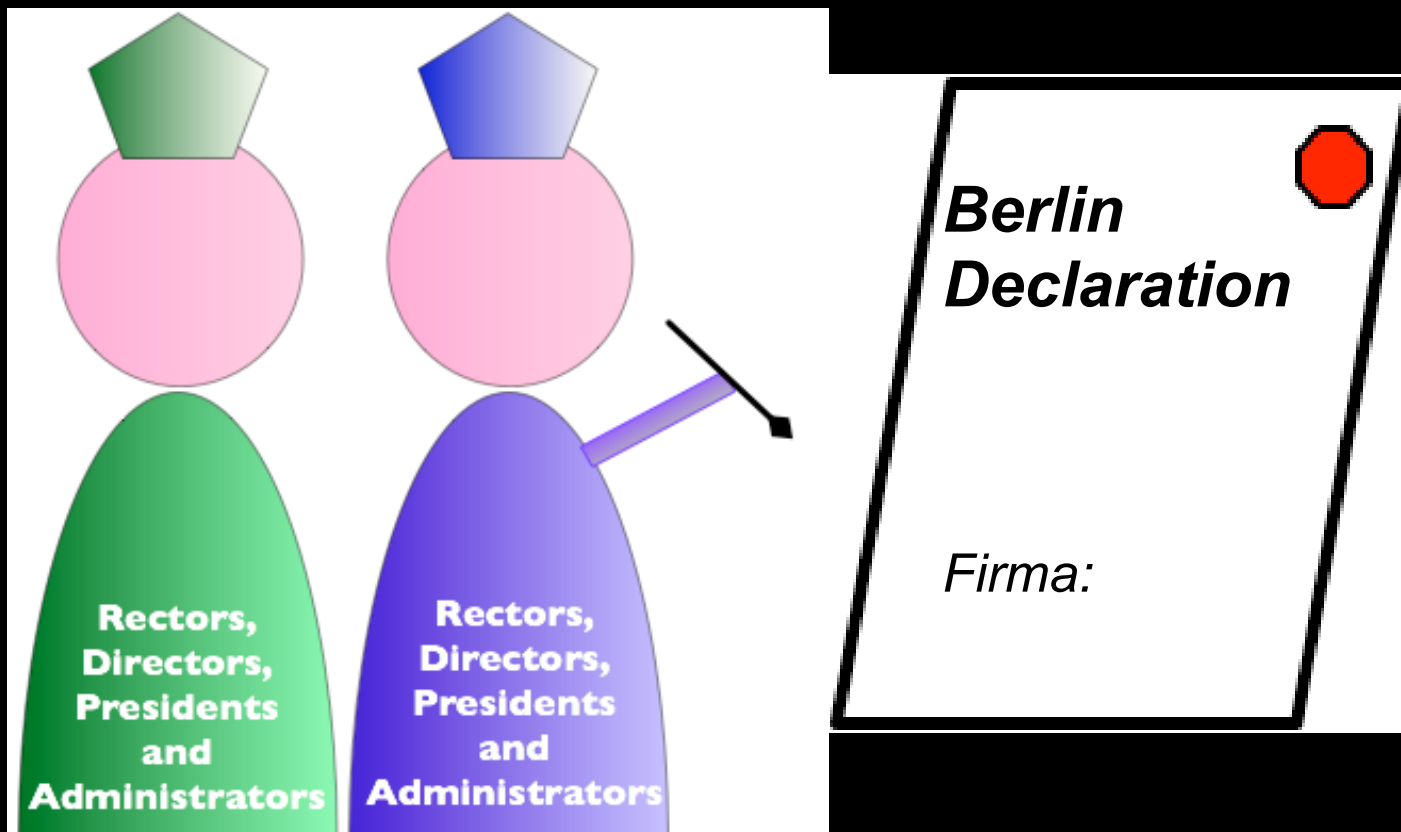
BUDAPEST OPEN ACCESS INITIATIVE (2002)

“We invite governments, universities, libraries, journal editors, publishers, foundations, learned societies, professional associations, and individual scholars who share our vision to join us in the **task of removing the barriers to open access and building a future in which research and education in every part of the world are that much more free to flourish.**”

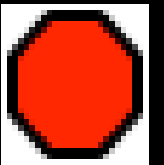
2. Storia > le istituzioni per l'open access

2003:

Un primo impegno istituzionale
per l'open access



**“BERLIN 3”:
Marzo 2005:**



Agreed Recommendation:

"In order to implement the Berlin Declaration institutions should implement a policy to:

1. **require** their researchers to **deposit a copy** of all their published articles in an **open access repository**
and
2. **encourage** their researchers to publish their research articles in open access journals where a suitable journal exists (and provide the support to enable that to happen)."

2. Storia > i numeri dell'open access

- Più di 150 istituzioni di ricerca hanno sottoscritto la Dichiarazione di Berlino
- 74 rettori di università italiane su 77 sono tra questi
- Più di 270 repository sono OAI-PMH compliant
- Più di 2100 riviste sono OA

Abbiamo a disposizione moltissimi dati e metadati
e
un patto condiviso su come usarli

Indice

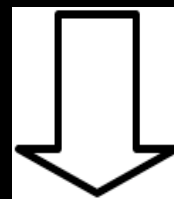
- 1. Definizioni introduttive
- 2. Storia
- 3. Filosofia
- 4. Strumenti
- 5. Problemi e prospettive
- 6. Link e bibliografia

3. Filosofia > idee e principi alla base del movimento

- **“Obiettivi (Dichiarazione di Berlino)**

La nostra missione di disseminazione della conoscenza è incompleta se l'informazione non è resa largamente prontamente disponibile alla società. Occorre sostenere nuove possibilità di disseminazione della conoscenza, solo attraverso le modalità tradizionali ma anche e sempre più attraverso il paradigma dell'accesso aperto via Internet. Definiamo l'accesso aperto come una fonte estesa del sapere umano e del patrimonio culturale che siano stati validati dalla comunità scientifica.

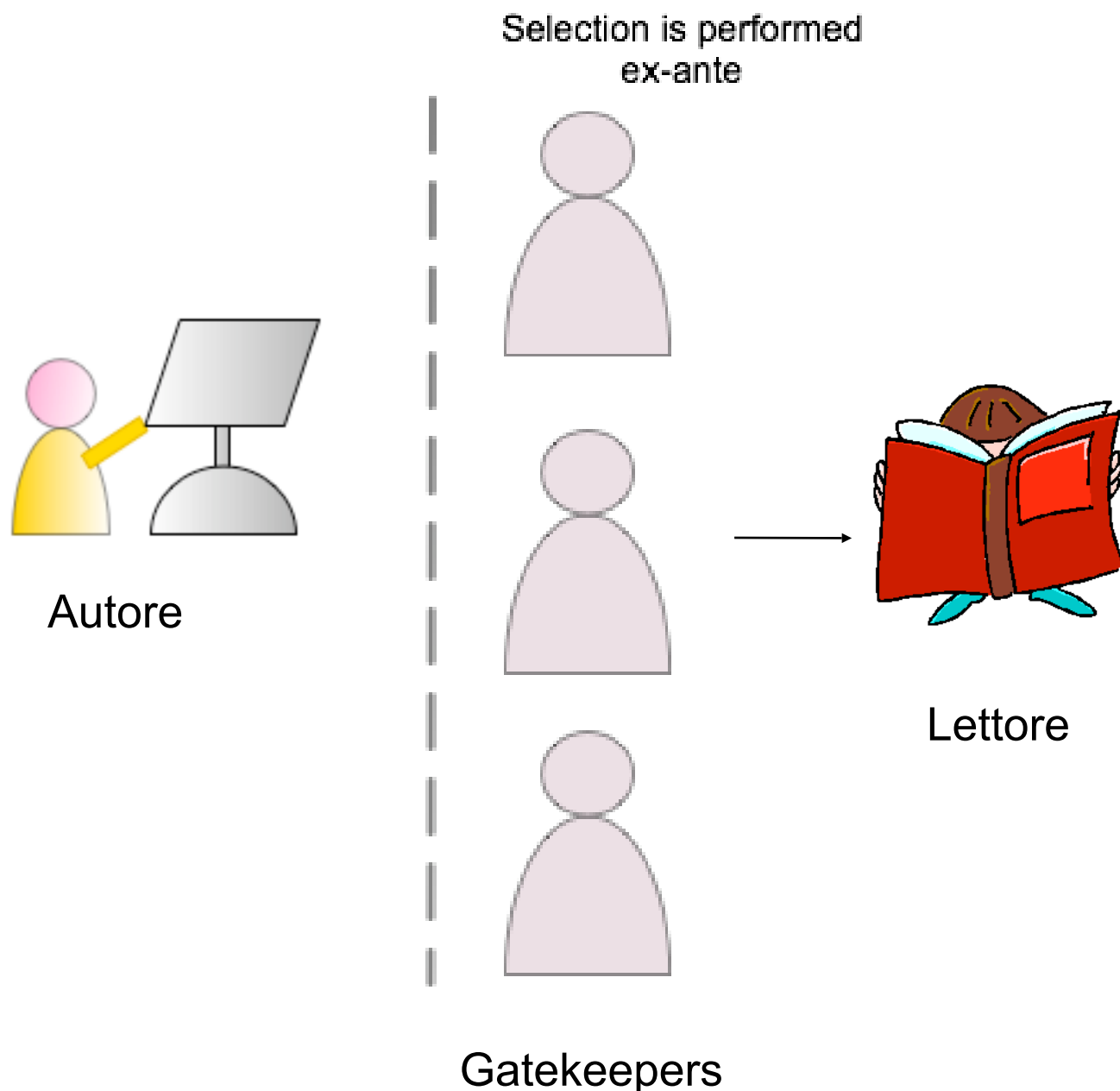
Per mettere in pratica la visione di un'istanza globale ed accessibile del sapere, **il Web del futuro dovrà essere sostenibile, interattivo e trasparente. I contenuti ed i mezzi di fruizione (tools) dovranno essere compatibili e ad accesso aperto.”**



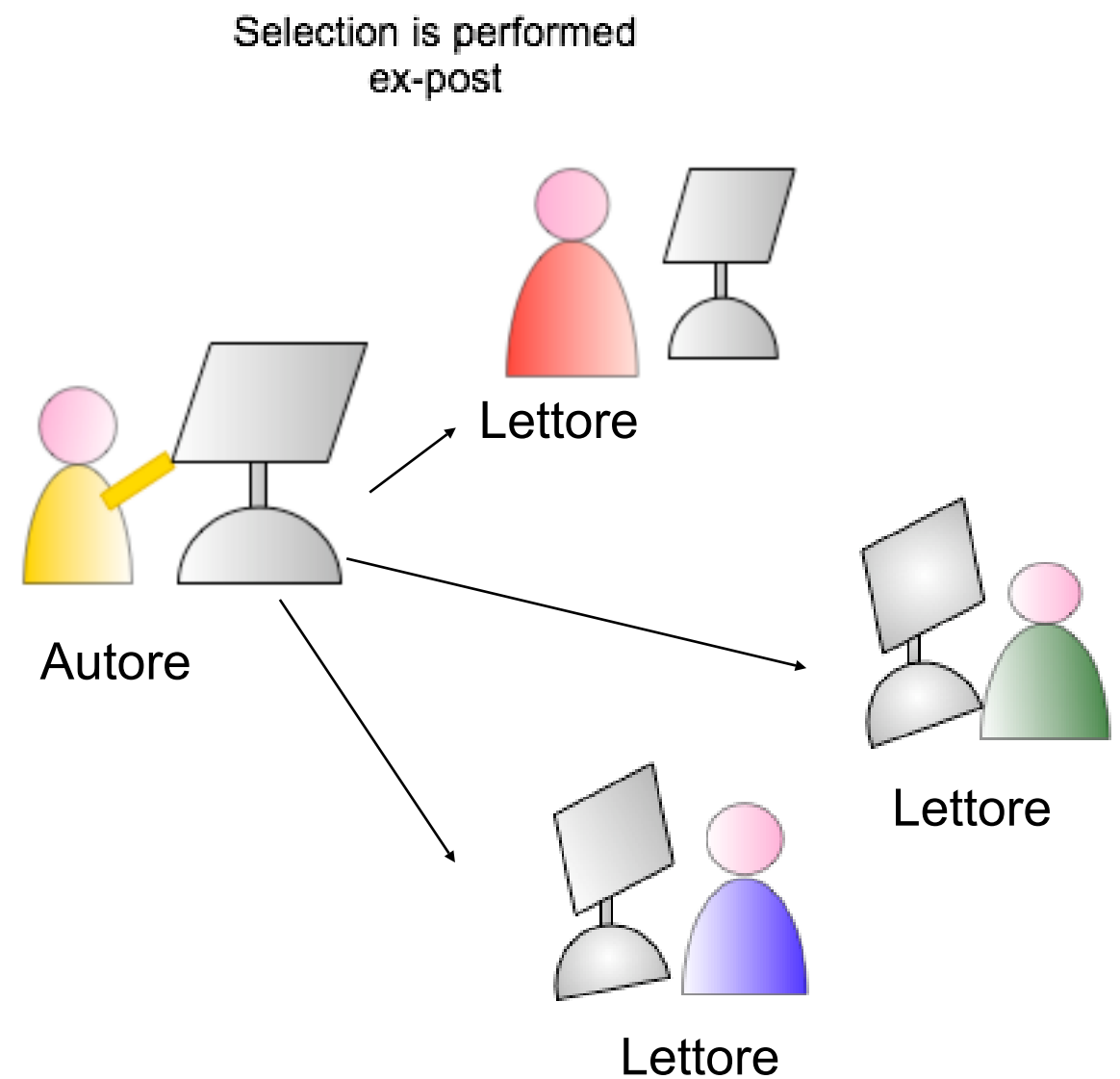
- * il sapere è scientifico e liberamente accessibile
(verificabile, emendabile)
- * Internet e Web: infrastruttura e “filosofia”: collegamento
 - * Selezione: ex post

3. Filosofia > 2 modelli di selezione

Selection before Access



Universality of Access



Indice

- 1. Definizioni introduttive
- 2. Storia
- 3. Filosofia
- 4. Strumenti
- 5. Problemi e prospettive
- 6. Link e bibliografia

4. Strumenti

Protocolli e standard: OAI-PMH e Dublin Core

Leggi e policies:

- Copyright: in Italia, Legge 22 aprile 1941, n. 633
- Licenze (CC)
- Editori e Open access: Romeo e Sherpa
- Il ruolo delle università: esempi

Software:

Archivi

- Eprints (esempio)
- DSPace
- CDSWare
- Fedora


Riviste

- OJS
- HyperJournal
-

4. Strumenti > Creative commons (e science commons)


scegli la licenza

Scegliendo una licenza Creative Commons mantieni i tuoi diritti d'autore ma permetti agli altri di copiare e distribuire la tua opera purché loro te ne riconoscano la paternità – e solamente sotto le condizioni che tu specifichi qui. Se vuoi offrire la tua opera senza alcuna restrizione all'uso scegli la donazione al pubblico dominio.


Permetti che la tua opera venga utilizzata a scopi commerciali? ([maggiori informazioni](#) )

☒ Sì


☐ No

Permetti che la tua opera venga modificata? ([maggiori informazioni](#) )

☒ Sì

☐ Sì, purché le opere derivate siano distribuite sotto la stessa licenza ([maggiori informazioni](#) )

☐ No

Giurisdizione della tua licenza ([maggiori informazioni](#) )

Italia 

Seleziona il formato della tua opera:

Altro 

[Inserisci maggiori informazioni sulla tua opera.](#)

Scegli una licenza

Creative commons



Attribuzione - Condividi allo stesso modo 2.0 Italia

Tu sei libero:

- di riprodurre, distribuire, comunicare al pubblico, esporre in pubblico, rappresentare, eseguire o recitare l'opera
- di creare opere derivate
- di usare l'opera a fini commerciali

Alle seguenti condizioni:



Attribuzione. Devi riconoscere il contributo dell'autore originario.



Condividi allo stesso modo. Se alteri, trasformi o sviluppi quest'opera, puoi distribuire l'opera risultante solo per mezzo di una licenza identica a questa.

- In occasione di ogni atto di riutilizzo o distribuzione, devi chiarire agli altri i termini della licenza di quest'opera.
- Se ottieni il permesso dal titolare del diritto d'autore, è possibile rinunciare ad ognuna di queste condizioni.

Le tue utilizzazioni libere e gli altri diritti non sono in nessun modo limitati da quanto sopra

4. Strumenti > le politiche degli editori

Project RoMEO

http://www.lboro.ac.uk/departments/ls/disresearch/romeo/

Getting Started Latest Headlines www.sp.unipi.it - Ho... PhpCollab Main Page - HJ_econ... Mantis Conoscenza e scienz...

Disable CSS Forms Images Information Miscellaneous Outline Resize Tools View Source Options

MIT OpenCourseW... Repubblica.it » Verso ... Project RoMEO Publisher copyright p... Publisher copyright p...



Project RoMEO

"Classical textual copyright law is designed for on-paper text that is sold for revenue. The primary protections on which copyright focuses concern protection of the author and the publisher from theft-of-copies-of-the-text, and only secondarily with other concerns, such as plagiarism, corruption, or misuse of the text (other than stealing it or selling it illicitly). In the digital era, and especially for texts from which the author does not seek sales revenue, the other rights must be very explicitly formulated. Professor's Oppenheim's is accordingly a very important and timely project that has my fullest support." *Professor Stevan Harnad*

The RoMEO Project (Rights METadata for Open archiving) is funded by the Joint Information Systems Committee for one year (1 August 2002 - 31 July 2003) to investigate the rights issues surrounding the 'self-archiving' of research in the UK academic community under the Open Archive Initiative's Protocol for Metadata Harvesting.

It will perform a series of stakeholder surveys to ascertain how 'give-away' research literature (and metadata) is used, and how it should be protected. Building on existing schemas and vocabularies (such as Open Digital Rights Language) a series of rights elements will be developed. A solution for the protection of the IPR in metadata itself will also be created.

→ **About RoMEO** → **Surveys**

→ **Project Team** → **Deliverables**

Romeo and Juliet, 1884
by Sir Frank Dicksee (1853-1928)
Southampton City Art Gallery,

CC BY SA

4. Strumenti > le politiche degli editori

Browser window: Publisher copyright policies & self-archiving - SHERPA

Address bar: <http://www.sherpa.ac.uk/romeo.php>

Navigation bar: Getting Started | Latest Headlines | www.sp.unipi.it - Ho... | PhpCollab | Main Page - HJ_econ... | Mantis | Conoscenza e scienz...

Tools bar: Disable | CSS | Forms | Images | Information | Miscellaneous | Outline | Resize | Tools | View Source | Options

Search bar: MIT OpenCourseW... | Repubblica.it » Verso ... | Project RoMEO | Publisher copyright p... | Publisher copyright p...

SHERPA

...opening access to research

[Home](#) | [About](#) | [Advice](#) | [Archives](#) | [Links](#) | [Contacts](#)

Publisher copyright policies & self-archiving: the SHERPA/ROMEEO list

Use this site to find a summary of permissions that are normally given as part of each publisher's copyright transfer agreement.

Search

Search for

find ☐ contains or ☒ starts with or ☐ Exact phrase only


[Or view all publishers](#)

This service is a development of the original publishers' listings produced by the [ROMEEO project](#).
Journal information is kindly provided by the British Library's [Zetoc](#) service hosted by MIMAS.

Publisher information is updated by community contributions - please submit any updates using the link provided on the search results pages.
If a publisher is not listed here you may wish to [recommend that we include it](#).

4. Strumenti > LOCKSS

[LOCKSS](#) [CLOCKSS](#) [PLNs](#) [About Us](#) [Contact Us](#)



HOME

Welcome

What is the LOCKSS Program?

LOCKSS (Lots of Copies Keep Stuff Safe), based at Stanford University Libraries, is an international community initiative that provides libraries with digital preservation tools and support so that they can easily and inexpensively collect and preserve their own copies of authorized e-content. LOCKSS, in its tenth year, provides libraries with the open-source software and support to preserve today's web-published materials for tomorrow's readers while building their own collections and acquiring a copy of the assets they pay for, instead of simply leasing them.

The ACM award-winning LOCKSS technology is an open source, peer-to-peer, decentralized digital preservation infrastructure. LOCKSS preserves all formats and genres of web-published content. The intellectual content, which includes the historical context (the look and feel), is preserved. LOCKSS is OAIS-compliant; the software migrates content forward in time [↻](#); and the bits and bytes are continually audited and repaired [🔧](#).


Today LOCKSS is a thriving international community-based initiative with libraries and publishers working together with the shared goal to preserve e-content for the long-term. More than 300 leading scholarly publishers have granted permission for their content to be preserved by LOCKSS Alliance members.

Why should you care?

NAVIGATION

- [Home](#)
- [Participating Libraries](#)
- [Publishers and Titles](#)
- [How It Works](#)
- [For Libraries](#)
- [For Publishers](#)
- [LOCKSS Alliance](#)
- [Talks](#)
- [Publications](#)
- [Installing LOCKSS](#)

“...let us save what remains:
not by vaults and locks
which fence them from the
public eye and use in
consigning them to the
waste of time, but by such a
multiplication of copies, as
shall place them beyond the

 SA

4. Strumenti >

le politiche di università (e UE)

“Best practices”: alcuni esempi

MIT OpenCourseWare DFG:

CERN: Geneva, 31 March 2005.

CERN confirms its commitment to open access to scientific information. At a meeting last Wednesday, the Organization's executive committee endorsed a policy of open access to all the laboratory's results, as expressed in the document 'Continuing CERN action on Open Access' (pdf), released by its Scientific Information Policy Board (SIPB) earlier in the month. "This underlines CERN's commitment to sharing the excitement of fundamental research with as wide an audience as possible", said Guido Altarelli, current SIPB chairman.

In Italia...

Università di Trento
(cfr. F.Valentini)

boaforum messages

[BOAI] DFG Passes Open Access Guidelines

From: Peter Suber <peters AT earlham.edu>

Date: Tue, 07 Mar 2006 22:38:02 -0500

DFG Passes Open Access Guidelines

Information for Researchers No. 04 30 January 2006

In 2003 the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) signed the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. The DFG supports the culture of open access. Unhindered access to publications increases the distribution of scientific knowledge, thereby enhancing the authors' visibility and contributing to their reputations.

The DFG has now tied open access into its funding policy. During their meetings in January 2006, the DFG's Senate and Joint Committee recommended encouraging funded scientists to also digitally publish their results and make them available via open access.

Further information on open access is available at
<<http://www.dfg.de/lis/openaccess>>www.dfg.de/lis/openaccess.

4. Strumenti > software (free/open source!!)

Archivi

(compatibili con OAI-PMH)

- Eprints Free Software (Southampton)
- DSpace (MIT, Boston + HP)
- CDSWare (CERN, Geneve)
- Fedora (Cornell University, Usa)

Riviste

(compatibili con OAI-PMH)

- Open Journal Systems - OJS (British Columbia University, Canada)
- HyperJournal (Pisa, Università e Associazione no-profit)
- GAP - German Academic Publishers (Hamburg)
- Lodel CMS - Logiciel d'édition électronique - Online publishing tool (Paris)

4. Strumenti > EPrints (www.eprints.org)

EPrints Free Software



[HOME](#) [OPEN ACCESS](#) [SOFTWARE](#) [COMMUNITY](#) [SERVICES](#)



Hi, my name is Christopher Gutteridge and I am the lead developer for EPrints. [Meet my team](#)

EPrints is both a practical tool and the crystallization of a philosophy. It enables research to be accessible to all, and provides the foundation for all academic institutions to create their own research repositories.

EPrints is a great piece of software but we have plans to make it better still. With the new EPrints Community initiative we will be able to work closely with new and existing users to address their needs and concerns, and allow them to help set our priorities. Working towards a more modular design will

Demonstration Sites

Our [demonstration site](#) allows you to try the features of our EPrints software. Detailed help is provided.

You can also see how EPrints users worldwide have [configured their repositories to meet their diverse requirements](#).

Support

Premium Support Services

For premium support and training, see [EPrints services](#).

Announcements Mailing List

This list is used to announce new versions of the EPrints software, and major news. It is not a discussion list.

To subscribe to the list, send a mail with one of the following lines in the message *body* to majordomo@ecs.soton.ac.uk

subscribe eprints-announce

Technical Mailing List


Quick Links:

- [Software Homepage](#)
- [Demo Site](#)
- [Download](#)
- [Sites Using EPrints](#)
- [Wiki](#)
- [Tech-List Archive](#)
- [Documentation](#)
- [Prerequisites](#)

4. Strumenti > Hyperjournal

(www.hjournal.org)

**HyperJournal**
Web Site



Home | Documentation | Downloads | Links | Overview

Menu

- Features
- ▣ News
- ▣ Roadmap
- Team and Partners
- ▣ Books
- Documentation
- ▣ Training
- ▣ Demo
- ▣ Download
- ▣ Mailing List
- ▣ Presentations
- Related Links
- ▣ How to Help
- ▣ Bugs and Suggestions
- ▣ Browse SVN Repository
- ▣ Commercial services
- ▣ Developers Wiki
- ▣ Subprojects

Hyperjournal

Open Access made Easy!

The HyperJournal is an Open Source software application which enables on-line as well as printed publishing in an innovative and significantly cost-cutting way. The HyperJournal can be easily installed, configured and managed without any out-of-the-ordinary IT-competence.

(categories: Home)

HyperJournal 0.5b Released

Submitted by barbz on March 22, 2006 - 16:45.

HyperJournal 0.5b (Beta) codename "Snowdomes" has been released.

This version features:

Donations

To make a donation just click the button below. You can donate via paypal and credit card. Thank you!!!

[Make a Donation](#)

Ads

Ads by Goooooogle

ClinPlus by DZS Software
Clinical Trials
Software for Data
Management,
Analysis &
Reporting
www.ClinPlus.com

International bidding

Find:

Altri strumenti

Internet archive (**wayback machine**):
<http://www.archive.org/index.php>

Ancora fravia:
http://www.searchlores.org/deepweb_searching.htm

Indice

- 1. Definizioni introduttive
- 2. Storia
- 3. Filosofia
- 4. Strumenti
- 5. Problemi e prospettive
- 6. Link e bibliografia

5a. Problemi

1. sostenibilità economica

e “selezione” (anche gestione delle carriere...)

2. SCI e IMPACT FACTOR (**IF**): Forma di misura standardizzata introdotta dall'Institute of Scientific Information (ISI). Si calcola su un ristretto insieme di riviste (**core journals**) dividendo il numero di citazioni che **una rivista** riceve, in un dato anno, relative agli articoli pubblicati nei due anni precedenti, per il numero di articoli pubblicati in quegli stessi anni.

3. trovare l'informazione scientificamente rilevante:

- nuovi strumenti di ricerca
- trasportare i dati, oltre che i metadati

5.b Prospettive

1. sostenibilità economica: modelli ibridi, riallocazione delle risorse, ma soprattutto prestigio e.. coraggio:-)!

e “selezione” (anche gestione delle carriere...)

2. Alternative all'Impact Factor: es. Misure dell'impatto basate su metriche di “Social Network Analysis”: permettono di ottenere risultati più significativi del semplice **IF**; considerano la posizione dell'articolo nel network, invece si basarsi solo sulle citazioni ricevute. Es. combinare citazioni e download; es. combinare popolarità e prestigio.

5b. Prospettive

3. trovare l'informazione scientificamente rilevante:

- **OpenDOAR - the Directory of Open Access Repositories.**

OpenDOAR è il registro degli open archive dell'università di Lund. Il progetto invita a registrare il proprio archivio

- **Registry of Open Access Repositories (ROAR).**

Anche ROAR è un registro degli archivi open access, mantenuto da Tim Brody (Southampton)

- **Directory of Open Access Journals (DOAJ).**

La directory dell'università di Lund contiene oltre 2100 riviste open access



- **strumenti di ricerca semantica:** es. estrattori di semantica nascosta; web services per ricostruire i percorsi genetici delle idee e dei testi.
- **trasportare i dati, oltre che i metadati:** es. integrazione con RSS 1.0;

5. prospettive > es. PubMed



PubMed Central e servizi “aggiunti”:

- es. Entrez
- es. HubMed: An alternative interface to the PubMed medical literature database

es.

- Quick access to searches with a Firefox search plugin or a HubMed bookmarklet (drag to your browser's bookmarks toolbar).

- Export citations in RIS, BibTeX, RDF and MODS formats, or directly to RefWorks.

- Graph occurrences of keywords in published papers over time.

- Tag and store annotated metadata for articles of interest.

- es. CiteULike

grazie